

Virtual patients for learning of clinical reasoning

Citation for published version (APA):

Huwendiek, S. E. (2016). *Virtual patients for learning of clinical reasoning*. [Doctoral Thesis, Maastricht University]. Datawyse / Universitaire Pers Maastricht. <https://doi.org/10.26481/dis.20161125sh>

Document status and date:

Published: 01/01/2016

DOI:

[10.26481/dis.20161125sh](https://doi.org/10.26481/dis.20161125sh)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

The diagram illustrates the diagnostic reasoning process, which is a cyclical and iterative process. It is represented by a large circle containing a flowchart of the process, with four orange arrows pointing from the process to the video consultation window below.

Diagnostic Reasoning Process:

- Patient's story** leads to **Information gathering**.
- Information gathering** leads to **Problem representation**.
- Problem representation** leads to **Pattern processor**.
- Pattern processor** leads to **Evaluation**.
- Evaluation** leads to **Calibration**.
- Calibration** leads to **Diagnosis**.

The process is supported by two systems:

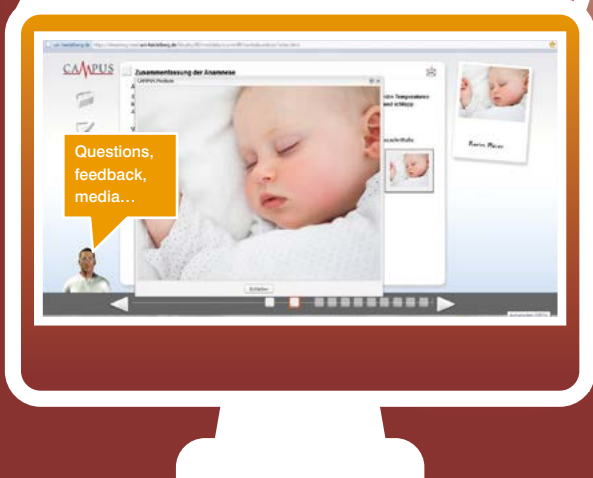
- System 1 processes «pattern recognition»** (left side of the circle).
- System 2 processes «analytical»** (right side of the circle).

The process is also influenced by **Context** (indicated by an arrow pointing to the top left of the circle).

Video Consultation Window:

The window shows a video of a baby. A speech bubble from the baby says: "Questions, feedback, media...". The window also displays a list of images and a small video of a person.

Sören Huwendiek



The research reported here was carried out through



within the School of Health Professions Education



at the Department of Paediatrics 1, University Children's Hospital, Heidelberg, Germany

ISBN 978 94 6159 617 8

Copyright S E Huwendiek, Weil am Rhein 2016

Layout and print: Datawyse | Universitaire Pers Maastricht



Virtual Patients for learning of clinical reasoning

DISSERTATION

to obtain the degree of Doctor at Maastricht University,
on the authority of the Rector Magnificus Prof. Dr. Rianne M. Letschert,
in accordance with the decision of the Board of Deans,
to be defended in public on Friday 25th November 2016, at 12.00 hours

by

Sören Eric Huwendiek

Supervisors

Prof. Dr. D.H.J.M. Dolmans

Prof. Dr. C.P.M van der Vleuten

Co-supervisor

Dr. B. de Leng

Assessment committee

Prof. Dr. J.J.E. Rethans (Chairman)

Dr. G. Clarebout

Prof. Dr. I.C. Heyligers

Prof. Dr. A.D.C. Jaarsma (UMC Groningen)

Dr.ir. P.G.M de Jong (UMC Leiden)

To my wife, Susanne, and son, Elias,
and my parents, Erika & Volker

CONTENTS

CHAPTER 1	General Introduction	9
CHAPTER 2	Establishing guidance for Virtual Patient design <i>Design principles for virtual patients: a focus group study among students</i> <i>Medical Education 2009; 43: 580–588</i>	25
CHAPTER 3	Typing Virtual Patient Design <i>Towards a typology of virtual patients</i> <i>Medical Teacher 2009; 31: 743-748</i>	43
CHAPTER 4	Evaluating Virtual Patient design <i>Exploring the validity and reliability of a questionnaire for evaluating virtual patient design with a special emphasis on fostering clinical reasoning</i> <i>Medical Teacher 2014; Oct 14:1-8. [Epub ahead of print]</i>	59
CHAPTER 5	Establishing guidance for Virtual Patients` implementation <i>Learner preferences regarding integrating, sequencing and aligning virtual patients with other activities in the undergraduate medical curriculum: A focus group study</i> <i>Medical Teacher 2013; 35: 920–929</i>	77
CHAPTER 6	Assessing Virtual Patients` implementation <i>Electronic assessment of clinical reasoning in clerkships: A mixed-method comparison of long menu key feature problems with context rich single-best-answer questions</i> <i>Manuscript under review</i>	99
CHAPTER 7	General discussion	123
	Summary	135
	Samenvatting (summary in Dutch)	141
	Valorization	149
	Acknowledgements	159
	Curriculum Vitae	163
	She dissertation series	167

CHAPTER 1

General Introduction

This chapter introduces the main concepts of this dissertation, discusses its background, and presents the dissertation's rationale, aim and research questions. Further, it outlines how each chapter contributes to the dissertation's aim.

BACKGROUND

What is clinical reasoning?

A four-week-old baby boy is brought to the emergency room by his mother at 03:00 a.m. in the morning because of repetitive vomiting. The vomiting had started suddenly at 00:45 a.m. after feeding and is forceful like a projectile. Since then he has been vomiting again and again after feeding. Yesterday he had his last bowel movements which were normal in consistency and colour. The last time he wet his nappy was yesterday evening. He is thirsty and is breast-fed by his mother. The pregnancy, birth and first three weeks of life were normal. The mother and father are healthy and have no diarrhoea or vomiting. He is the first child of the family. In the physical exam you find a child that looks a little emaciated. You find a palpable 'olive-like' mass in the right upper quadrant of the abdomen, but otherwise the physical exam is normal.

What is the probable diagnosis? To find the most likely diagnosis one needs to look at each piece of available information and search for the most plausible explanation (Eva 2014). Such task is rather complex: It often entails attentive observation, appropriate elicitation of the medical history, correct performance of physical examinations, the generation of hypotheses, appreciation of the relationship between each piece of data and each hypothesis, and attempts to confirm / disconfirm hypotheses through the appropriate ordering of diagnostic tests (Eva 2004). These processes are called 'clinical reasoning', which is described as a 'context-dependent way of thinking and decision making in professional practice to guide practice actions' (Higgs et al. 2008). In other words, clinical reasoning involves both the establishment of a diagnosis and the decision on a management plan that is specific to the patient's circumstances and preferences (Durning et al. 2011). Figure 1 (page 11) presents a model of clinical reasoning which integrates key features of several theoretical frameworks. The main elements that make up this model are the following, as described by Gruppen and Frohna [2002], Bowen 2006, Croskerry 2009 and summarized in Huwendiek et al. 2014:

- The patient's story: This is all about the patient, including all of his/her symptoms and characteristics.
- Context: This refers to the situational factors which may have an important bearing on clinical reasoning processes.
- Knowledge and experience: The abilities and skills the physician has accumulated so far and which are key co-determinants of clinical reasoning outcomes.

- **Information gathering:** During this process, the physician interviews the patient to obtain the information necessary to reach the correct diagnosis.
- **Problem representation:** Drawing on his own expertise gained so far, the physician decides what information elements to take in and which ones to ignore, thereby creating his/her own mental image of the patient. The clinical reasoning process is one of continuous interpretation and reinterpretation.
- **Evaluation:** once a first mental image of the patient has been created, the physician constantly revisits his/her preliminary interpretation, setting it against situational factors and the patient's characteristics that have come to light in the course of the encounter. This outcome may lead the physician to ask additional questions that allow him/her to make assumptions and finally arrive at the correct diagnosis. These hypotheses come to the mind of the physician because cues in the patient's history activate 'illness scripts' from the physician's memory, or mental representations of a typical patient with a specific disease (Schmidt et al. 1990, Schmidt & Mamede 2015). These scripts contain findings that are expected to be present in patients with that disease and guide physicians in deciding which information needs to be gathered through history taking, a physical exam or diagnostic tests (Schmidt & Mamede 2015). In general this evaluation process can be rather non-analytical, when large amounts of cases created a sense of pattern recognition of symptoms (System 1 processes), or rather analytical, when a person deliberately gathers information for rational decisions (System 2 processes) (Norman, 2005; Croskerry 2009). Depending mainly on the complexity of a clinical case and the experience of the physician, either the one or the other mode is used (Eva 2004; Croskerry 2009). A more detailed overview of these processes is presented in Croskerry's universal model of clinical reasoning (Croskerry 2009).

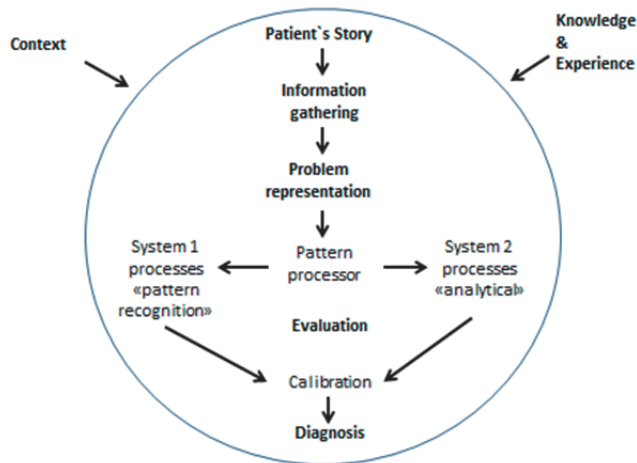


Figure 1: A model of clinical reasoning integrating key features of several theoretical frameworks (adapted from Bowen 2006, Croskerry 2009 and Gruppen & Frohna [2002])

Why is clinical reasoning and the teaching of it important?

Clinical reasoning is generally considered a decisive component of nearly everything doctors do in practice (Higgs et al. 2008, Norman 2005, Eva 2005, Durning et al. 2011). Clinical reasoning enables doctors to establish the right diagnosis and make the right treatment decisions for a patient (Norman 2005, Eva 2005). Croskerry put it in the following words: 'In fact, it is difficult to imagine anything of greater importance or relevance to patient outcomes and to patient safety' [than clinical reasoning] (Croskerry 2009).

However, doctors do not make correct diagnoses or treatment decisions as often as generally anticipated: Overall, the diagnostic failure rate is estimated to be between 10 and 15%, with the highest rates among specialties in which patients are diagnostically undifferentiated, such as emergency medicine or family medicine, and the lowest rates in visual specialties, such as radiology and pathology, with an estimated rate of about 2% (Berner & Graber 2008, Croskerry 2013). Because clinical reasoning plays such a major role in final patient outcomes, acquiring this competence is considered essential for a physician.

The good news is that it can be learned and improved. In general it has been shown that students develop clinical reasoning skills by seeing many patients, actively engaging in problem-solving and receiving sufficient feedback (Eva 2005, Norman 2005). Strategies have been described to specifically support the development of clinical reasoning when seeing real patients, which can be used by the teacher to foster learning of clinical reasoning (Bordage 1999, Nendaz & Bordage 2002, Bowen 2006). These strategies are displayed in Table 1, page 16, first section.

However, medical education seems to fall short of providing students with adequate opportunities to learn these skills, exposing them to too few and less varied real patients (Rattner et al. 2001), and of providing adequate feedback and coaching (Wimmers et al. 2005, Schmidt & Mamede 2015, Eva 2004). An alternative to using real patients to foster learning of clinical reasoning is to use Virtual Patients (Cook & Triola 2009), which will be defined in the following paragraph.

What are Virtual Patients?

Virtual Patients (VP) are online cases in which the learner takes on the role of physician, has to make all decisions him/herself and receives feedback on these decisions (Cook & Triola 2009, Huwendiek & Haag 2012). They have been defined as 'interactive computer simulation[s] of real-life clinical scenarios for the purpose of medical training, education, or assessment' (Ellaway et al. 2008). It is important not to confuse VPs with other educational instruments that are sometimes mistakenly referred to as VPs but that in fact have little to do with the Virtual Patients under discussion here, such as: (i) 'simulated patients' or 'standardised patients', who are healthy persons acting like patients so that

learners can practise their communication skills, for example, and (ii) patient simulators in the form of manikins or parts thereof, to allow experiential learning of practical skills in an error-forgiving environment (Huwendiek et al. 2009b). The ability of VPs to foster clinical reasoning is considered as their most important asset (Cook & Triola 2009). Many different platforms and designs of Virtual Patients have been described (Fall et al. 2005, Bryce et al. 1998, Garde et al. 2005, Zary et al. 2006). A typical virtual patient session includes a stepwise unfolding of the case scenario by interviewing, examining, diagnosing and treating the electronic patient and receiving feedback on actions taken. Figure 2 presents a screenshot that includes an example of a VP (page 13). Why VPs are used in medical education will be addressed in the following paragraph.

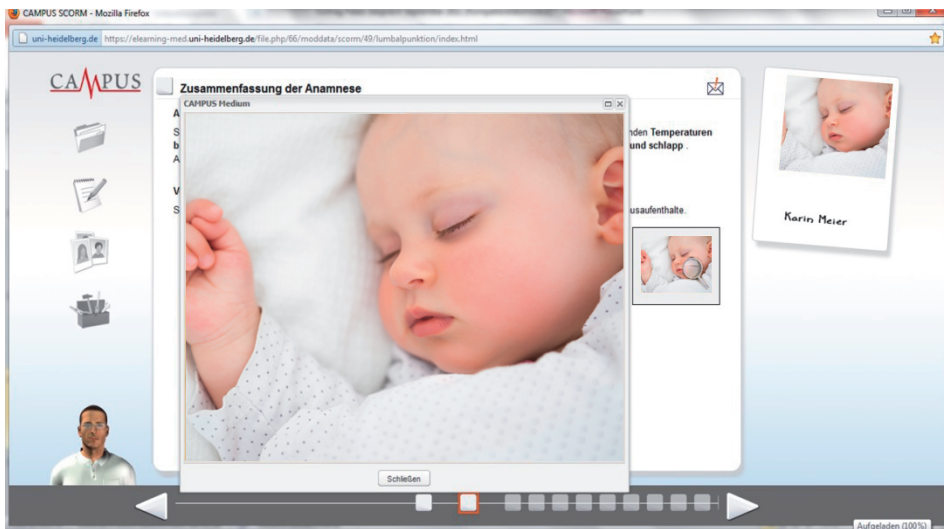


Figure 2: Screenshot of VP example

Why use Virtual Patients?

As previously stated, research has shown that students develop clinical reasoning skills by seeing many different patients, actively engaging in problem-solving and receiving sufficient feedback (Eva 2005, Norman 2005). However, there seem to be mainly two problematic aspects at play: In general, few real patients are actually suitable for teaching (changes in contemporary health care systems have caused patients to be hospitalised for shorter periods, for example), and supervisors often fail to encourage their students to involve heavily in clinical reasoning and to provide them with feedback (Irby 2004, Huang et al. 2007). The impact of Virtual Patients on students' learning outcomes has been studied and reviewed, and findings suggest that they can be effective especially for clinical reasoning (Botezatu et al. 2010; Cook et al. 2010; Cook & Triola 2009). In

these respects the use of VPs in conjunction with real patients for learning clinical reasoning offers several advantages.

First, VPs create a safe learning environment for students in which they learn by trial and error, while real patients remain shielded from students' mistakes. Second, they afford the opportunity to learn clinical reasoning when no suitable real patients are available. As Huang recapitulates, Virtual Patients 'fill gaps in clerkships by exposing students to diseases that they would not otherwise experience because of short clinical rotations and limited ambulatory care experiences' (Huang et al. 2007). This idea even made its way into the official documents of the accreditation body for medical programmes in the USA and Canada, the Liaison Committee for Medical Education: In its ED-2 standard they stated that students should encounter the types of patients and clinical conditions that are defined by the faculty. Gaps in this clinical exposure may be remedied by Virtual Patients or other simulated experiences (LCME, 2011). One typical example where VPs are especially suited are critically ill patients. Students are usually not heavily involved in the reasoning in such patients, as experts take care of these patients. Furthermore, there are also rare (but important) and seasonal diseases for which real patients are not always available. In such cases VPs can be used to train clinical reasoning. However, it cannot be stressed enough that VPs are not meant to replace real patients; rather, they are an *addition* to real patients to foster learning (Cook et al. 2010).

A third advantage is that many learners can be confronted with and learn from the same virtual patient cases. Already in 2011, for instance, more than 70 US and Canadian medical schools were using the same Paediatric VPs (Berman et al., 2011). This makes the development very cost-efficient. Imagine that a real clinical supervisor would have to repeatedly ask all the questions a virtual supervisor integrated into the VPs would normally ask students, this would be extremely resource-intensive. From this perspective it seems meaningful to pay close attention to the design of VPs and then expose many students to them. But how should VPs be designed? This will be addressed in the following paragraph.

Design of VPs to foster clinical reasoning: current approaches and challenges

What is known about VP design so far? The overarching design of Virtual Patients is modelled on a virtual patient-physician encounter, following a sequence of a patient presenting a complaint, patient interview, physical exam, diagnosis and management. VPs of the early 1960s already followed this design (de Dombal et al. 1969), which still presents the backbone of VP design today. Several studies on VPs have been conducted so far, which have even been summarised in a systematic review and meta-analysis (Cook et al. 2010). This review summarises 54 studies on VPs, 11 of which compared different designs of VPs. The authors suggest that learning outcomes may be improved if the design of VPs is based on repetition until demonstration of mastery, advanced organisers, enhanced feedback, and explicit contrasting of cases (Cook et al. 2010). Furthermore, there are general tips

available on how to develop VPs (Posel et al. 2009). Altogether, there are studies focusing on VP design, but the existing ones often focus on specific aspects other than clinical reasoning, such as communication skills (Bearman et al. 2001).

A first challenge, therefore, is to answer the question of how to specifically design a VP to foster the learning of clinical reasoning. For example, how to incorporate learner support in the form of a virtual supervisor within a VP to optimally facilitate the development of clinical reasoning? Which kind of questions should the virtual supervisor ask? In the absence of clear evidence-based guidance on how to design VPs to foster clinical reasoning this is a challenging task and it seems worthwhile to look at studies that do not address VP design itself but, nevertheless, can provide a theoretical framework for the design of such VPs. For the instructional design of VPs to foster clinical reasoning the following strategies and theoretical frameworks seem to be helpful: strategies to foster clinical reasoning (Bowen 2006, Nendaz & Bordage 2002; Bordage 1999, Schmidt & Mamede 2015), development principles for teaching cases (Kim et al. 2006), lessons from simulation research (Issenberg et al. 2005) and design principles for multimedia learning including cognitive load theory (Grunwald & Corsbie-Massay 2006, Stemler 1997, Mayer 2005, Lawless 1997, Merriënboer & Sweller 2010). For details and possible lessons learned about VP design, please refer to Table 1 on page 16. The table includes aspects that could be relevant when designing VPs to foster clinical reasoning.

Another challenge is the high variability in VP systems and designs (Bryce et al. 1998, Fall et al. 2005, Garde et al. 2005, Zary et al. 2006, Fischer 2000) that are often collectively labelled ‘Virtual Patients’, without specifying their design. To illustrate some differences: While some VP cases can have different outcomes as learners follow a different path depending on their actions, others unfold in a predetermined way after learners have received direct feedback. Furthermore, some VPs use only a few multiple choice questions (MCQs), while others use a combination of many MCQs and other question types. Some VPs use a lot of media including video while others do not make use of any visual media at all. On balance, the literature is packed with descriptions of VPs that often differ significantly in terms of instructional design as illustrated above, but at the same time these differences are often not clearly reflected and acknowledged in the respective scientific literature. As a result, misunderstandings and unwanted consequences such as conflicting study results may prevail, as apples might be compared with oranges. As language shapes and delimits our understanding, having a well-defined and common set of terms helps to identify and understand differences and similarities between different designs. A key step towards a better understanding and reporting of VP design would therefore be to achieve a consensus among VP experts on a well-defined and unambiguous language and concepts to describe Virtual Patient design, for instance in the form of a typology.

A third challenge is that little is known on how to evaluate VPs that are designed to foster clinical reasoning. Which aspects of VP design should such an instrument address? Should it ask about all design features (questions, feedback, media, etc.) or ra-

ther enquire into the effect these features might have on the learner? Such an evaluation instrument should be based on a theoretical framework and should be validated to allow for improved judgements and feedback in the process of developing VPs to foster clinical reasoning (Cook & Beckmann 2006). However, to our knowledge, no such instrument currently exists. As there is no theoretical framework on VP design to foster clinical reasoning readily available, the conceptual basis for the development of such an instrument has to be generated from existing frameworks and strategies. Especially helpful in this regard are the model of clinical reasoning presented in Figure 1 (see page 11) and the strategies for VP design to foster clinical reasoning described in Table 1 (page 16), from which we can derive the main concepts.

Table 1: Possibly helpful literature and frameworks for the design of VPs from outside of the existing VP literature
(informed by Huwendiek et al. 2009a)

Topic	Lessons learned	Literature
Strategies to foster clinical reasoning	<ul style="list-style-type: none"> • ask open-ended questions; • provide single-sentence summaries of patient problems in abstract terms; • ask for discriminating features of a set of diagnostic hypotheses; • probe for differential diagnoses in an early phase; prioritise diagnoses; • encourage students to compare and contrast diagnostic hypotheses based on real clinical data • demonstrate typical presentations of different diagnostic hypotheses and the relative probabilities of different diagnoses 	Bowen 2006; Nendaz & Bordage 2002; Bordage 1999; Schmidt & Mamede 2015
Core attributes for the development of teaching cases	Teaching cases should be: <ul style="list-style-type: none"> • relevant; • realistic; • engaging; • challenging; • instructional 	Kim et al. 2006
Lessons from simulation research	<ul style="list-style-type: none"> • provide feedback; • expose students to a wide range of difficulty levels; • involve multiple learning strategies; • capture clinical variation; • use them in a controlled environment; • provide for individualised learning and defined outcomes 	Issenberg et al. 2005
Design principles for multimedia learning, incl. cognitive load theory	<ul style="list-style-type: none"> • implement interactivity to promote active learning and critical thinking • synchronise audio and visual information; • eliminate multi-tasking; • maintain a stable learning environment; • eliminate redundant information; • decrease predetermined navigational control as expertise develops • provide authentic contexts 	Grunwald & Corsbie-Massay 2006; Stemler 1997; Mayer 2005; Lawless 1997; van Merriënboer & Sweller 2010

The above-mentioned three challenges, specifically the absence of concrete advice on how to design VPs to foster clinical reasoning, the lack of consensus on relevant VP design features and the missing of a validated VP design evaluation instrument, all have one central question in common which leads to the first overall research question of this dissertation:

How to improve the design of VPs to foster learning, especially of clinical reasoning?

Besides the design also the implementation of VPs seems relevant for fostering clinical reasoning, which will be addressed in the following section.

Implementation of VPs to foster clinical reasoning: current approaches and challenges

While it was previously believed that the benefits of digital tools such as VPs would be fully reaped just by introducing students to them, recent evidence has revealed that students are unlikely to embrace, apply and appreciate such tools so long as they are inadequately integrated with or even segregated from other parts of the curriculum (Haag et al. 2007; Fischer et al. 2008, Huwendiek et al. 2013). Regarding the implementation of VPs, the few available studies suggest that the following aspects are beneficial: VP-related student instructions, follow-up seminars, faculty development, formal assessment of VP content and avoidance of excessive workload by eliminating other teaching activities (Fischer et al. 2008, Berman et al. 2009, Edelbring et al. 2012).

More specifically, knowledge on how best to implement VPs to foster learning of clinical reasoning is scant. For example, which sequencing of VPs, real patient contacts and other educational events would be most beneficial? How should different educational activities be aligned with VP activities? These questions cannot be readily answered by the existing VP literature (Cook et al. 2010). Possibly helpful frameworks and strategies that can be found outside of the VP literature address the following aspects: the sequencing of educational activities (Hull et al. 2009, Ritter et al. 2007, Van Merriënboer & Sweller 2010, Issenberg et al. 2005), alignment of computer-mediated with other teaching activities (Bunk & Graham 2006, Choules 2007, Bonk 2006, Prober & Heath 2012, Prober & Khan 2013), alignment with assessment (McLachlan 2006, Cilliers et al. 2012, Biggs 1996, van der Vleuten & Schuwirth (2005)) and the development of clinical reasoning skills (Bowen 2006; Nendaz & Borge 2002; Borge 1999, Schmidt & Mamede 2015). For details and possible lessons learned for VP implementation, please refer to Table 2, page 19. The table includes aspects that could be relevant when implementing VPs to foster clinical reasoning.

When implementing VP activities, these should also be formally assessed because assessment drives learning (Cilliers et al. 2012, van der Vleuten & Schuwirth (2005)). Additionally, as the constructive alignment of learning objectives, teaching activities and assessment is regarded essential to support learning (Biggs 1996), the question of how

to assess VP activities is key to the implementation of VPs. Hence, the question is not **IF** learning from VPs should be assessed, but **HOW**. As this dissertation focuses on VPs whose intent is to foster learning of clinical reasoning, the question is which type of questions might be most appropriate to assess clinical reasoning. Different question formats for assessment of clinical reasoning in clerkships have been described (Higgs et al. 2008). Among the most commonly used question formats for assessing clinical reasoning are so-called Key-Feature Problems (KFPs) and Single Best Answer questions with a case vignette as stimulus format (context-rich SBAs, crSBAs) (Higgs et al. 2008). KFPs consist of a brief stem with a short patient vignette (stimulus format) followed by one or more questions. To date, studies into KFPs have found that the most typical KFPs with the best psychometric characteristics used a case vignette as stimulus format and short menu, a write-in or electronic Long-Menu response format and two to three questions per case/vignette (Bordage et al. 1995; Page & Bordage 1995, Fischer et al. 2005, Norman et al. 2006, Hrynychak et al. 2014, Bronander et al. 2014). Context-rich SBAs also have a case vignette and the response format mostly consists of 3 to 5 options with only one right answer. Usually only one question relates to the vignette and not more (Higgs et al. 2008). As single best answer question formats are viewed as more efficient, easier to score, more defensible and more cost-effective, they are often preferred to open-ended questions, such as write-in or electronic Long-Menu questions (LMQs) (Elstein 1993, Downing 2002, 2009, Desjardins et al. 2014). However, evidence suggests that the response format chosen may also affect learning, with cues influencing item difficulty and, consequently, open-ended questions being considered more difficult than closed-ended questions (Hemskerk et al. 2008, Desjardins et al. 2014). Such a perception may influence the educational effect (pre-assessment factors, Cilliers et al. 2012) of an exam, making it interesting to investigate whether the educational effect would be different when crSBAs or KFPs are used to assess VP activities and if so, how, and what the psychometric properties of such questions would be. To our knowledge, no study has ever compared the educational effects of using KFP and crSBA formats for the assessment of clinical reasoning in clinical clerkships.

The above-mentioned challenges concerning the implementation of VPs in general and the assessment of VP activities in particular, both aimed to foster clinical reasoning, have one central question in common which leads to the second overall research question of this dissertation:

How to improve the implementation of VPs to foster learning, especially of clinical reasoning?

Table 2: Helpful literature and frameworks for the implementation of VPs from outside of the existing VP literature
(informed by Huwendiek et al. 2013)

Topic	Lessons learned	Literature
Sequencing	<ul style="list-style-type: none"> • make sure the amount of novelty and amount to learn is maximised and adapted to learners' knowledge level; • consider the amount of time and repetitions required to learn something as well as the spacing of practice; • sequence learning tasks from simple to complex and from low-fidelity to high-fidelity environments 	Ritter et al. 2007; Van Merriënboer & Sweller 2010; Issenberg et al. 2005
Alignment of computer-mediated activities with other teaching activities	<ul style="list-style-type: none"> • strive to achieve the most advantageous combinations of face-to-face instruction and computer-mediated instruction to maximise the strengths and minimise the weaknesses of each instructional method 	Bunk & Graham 2006; Choules 2007; Bonk 2006; Prober & Heath 2012, Prober & Khan 2013
Alignment with assessment	<ul style="list-style-type: none"> • make sure to assess clinical reasoning, as it is generally accepted that assessment drives learning • align learning objectives, teaching and assessment in terms of both content and methods (constructive alignment). E.g.: when the learning objective is to be able to actively perform clinical reasoning, it should be actively learned and assessed 	McLachlan 2006; Cilliers et al. 2012; Biggs 1996 van der Vleuten & Schuwirth (2005)
Strategies to foster clinical reasoning	<ul style="list-style-type: none"> • For strategies designed to foster clinical reasoning, refer to Table 1, page 16 	Bowen 2006; Nendaz & Bordage 2002; Bordage 1999; Schmidt & Mamede 2015

RATIONALE, AIM AND MAIN RESEARCH QUESTIONS

The studies discussed in the foregoing indicate that Virtual Patients are especially suited to foster clinical reasoning (Botezatu et al. 2010; Cook et al. 2010; Cook & Triola 2009). However, there is a lack of evidence-based insights as to how to design and implement VPs to foster learning, especially of clinical reasoning (Cook et al. 2010).

The aim of this dissertation is therefore to explore how learning with Virtual Patients to foster clinical reasoning in medical students can be improved by addressing the following two main research questions:

- I: How to improve the design of Virtual Patients to foster learning, especially of clinical reasoning?
- II: How to improve the implementation of Virtual Patients to foster learning, especially of clinical reasoning?

DISSERTATION OUTLINE

Design of Virtual Patients to foster learning, especially of clinical reasoning

The first main research question is addressed in Chapters 2-4.

Chapter 2 focuses on the establishment of principles for the design of VPs to foster clinical reasoning, based on senior medical students' perceptions. This study used a qualitative approach after students had been exposed to VPs of a standard design and a design taking relevant frameworks from the literature into account. Students were interviewed about the impact different VP designs had on their learning of clinical reasoning and about what, in their view, would make the VP design that best fostered their learning, especially of clinical reasoning.

Chapter 3 describes the establishment of relevant features of VP design, based on perceptions of VP experts, and, consequently, how VP design should be reported in the scientific literature to improve educational research on the design of VPs. This study used the literature and a modified Delphi technique to establish a consensual typology, which was then validated by applying it to the description of different VP designs. The results from the study in Chapter 2 also informed the development of the typology.

Chapter 4 deals with the development and establishment of validity evidence for an instrument with which to evaluate VP design to foster clinical reasoning from the perspective of students. The questionnaire was developed by an international VP expert team and used a mixed-methods approach, also taking the results of Chapters 2 and 3 into account. Statistical validity evidence was established by using exploratory factor analysis and generalisability theory in a big sample of international students using VPs.

Implementation of Virtual Patients to foster learning, especially of clinical reasoning

The second main research question is addressed in Chapters 5-6.

Chapter 5 focuses on the establishment of guidance for the implementation of VPs using focus groups after senior medical students had been exposed to VPs in different curricular integration scenarios, which differed in the sequencing of and alignment between VPs and other educational formats, tutor involvement, relevance of VPs to assessment and the involvement of real patients. We explored students' perceptions of the scenarios and of VP implementation aspects that they thought would maximise their clinical reasoning skills.

Chapter 6 reports on how two different written exam formats in a summative assessment of a clinical clerkship are perceived by senior medical students, especially with respect to their influence on the learning of clinical reasoning (with VPs), and on whether their psychometric properties differ using a mixed-methods approach.

Chapter 7 discusses the findings reported in the various chapters and puts them into perspective, considers the implications for VP design and implementation and reflects on directions for future research and practical implications. A summary and valorisation of the results concludes the dissertation.

This dissertation is based on five studies on a related topic. Since every chapter was written with the intention to be read on its own, repetition and overlap between chapters is inevitable.

REFERENCES

- Bearman M, Cesnik B (2001): Comparing student attitudes to different models of the same virtual patient. *Medinfo* 2001;10:1004–8.
- Berman N, Fall LH, Smith S, Levine DA, Maloney CG, Potts M, Siegel B, Foster-Johnson L. (2009): Integration strategies for using virtual patients in clinical clerkships. *Acad Med* 84:942–949.
- Berner ES, Graber ML (2008): Overconfidence as a cause of diagnostic error in medicine. *Am J Med* 2008;121:Suppl:S2-S23.
- Biggs J (1996): Enhancing teaching through constructive alignment. *Higher Education* 32: 347-364, 1996. Kluwer Academic Publishers. Printed in the Netherlands.
- Bonk CJ, Graham CR. 2006. *Handbook of blended learning: Global perspectives, local designs*. San Francisco: Pfeiffer Publishing.
- Bowen JL. Educational strategies to promote clinical diagnostic reasoning. *N Engl J Med* 2006;355:2217–25.
- Bordage G, Carretier H, Bertrand R, Page G (1999): Comparing times and performances of French- and English-speaking candidates taking a national examination of clinical decision-making skills. *Acad Med* 1995;70:359–65.
- Bordage G. Why did I miss the diagnosis? Some cognitive explanations and educational implications. *Acad Med* 1999;74 (Suppl):138–43.
- Bronander KA, Lang VJ, Nixon LJ, Harrell HE, Kovach R, Hingle S, Berman N (2014): How we developed and piloted an electronic key features examination for the internal medicine clerkship based on a US national curriculum. *Med Teach*. 2014 Dec 11:1-6. [Epub ahead of print]
- Bryce D, King N, Graebner C, Myers J. Evaluation of a diagnostic reasoning program (DxR): exploring student perceptions and addressing faculty concerns. *J Interactive Media Educ*. <http://www.jime.open.ac.uk/98/1/bryce-98-1.pdf> [Accessed 10 September 2008].
- Bonk CJ, Graham CR. 2006. *Handbook of blended learning: Global perspectives, local designs*. San Francisco: Pfeiffer Publishing.
- Choules AP (2007): The use of elearning in medical education: a review of the current situation. *Postgrad Med J*. 2007 Apr;83(978):212-6.
- Cilliers FJ, Schuwirth LW, van der Vleuten CP (2012): A model of the preassessment learning effects of assessment is operational in an undergraduate clinical context. *BMC Med Educ* 12:9.
- Cook DA, Erwin PJ, Triola MM (2010): Computerized virtual patients in health professions education: a systematic review and meta-analysis. *Acad Med*. 2010 Oct;85(10):1589-602
- Croskerry 2009: A universal model of diagnostic reasoning. *Acad Med*. 2009 84(8):1022-8.
- Croskerry P (2013): From Mindless to Mindful Practice — Cognitive Bias and Clinical Decision Making. *N Engl J Med* 368;26
- de Dombal, F. T., Hartley, J. R., & Sleeman, D. H. (1969). A computer-assisted system for learning clinical diagnosis. *Lancet*, 1(7586), 145-148.
- Desjardins I, Touchie C, Pugh D, Wood TJ, Humphrey-Murto S (2014): The impact of cueing on written examinations of clinical decision making: a case study. *Med Educ*. 2014;48:255-61.
- Downing SM. Assessment of knowledge in written test forms. In: Norman G, van der Vleuten C, Newble D, eds. *International Handbook of Research on Medical Education*. Dordrecht, Netherlands: Kluwer Academic 2002;642–72.
- Downing SM. Written tests: constructed response and selected response formats. In: Downing SM, Yudkowsky R, eds. *Assessment in Health Professions Education*. New York, NY: Routledge 2009;149–84.
- Durning S, Artino AR Jr, Pangaro L, van der Vleuten CP, Schuwirth L (2011): Context and clinical reasoning: understanding the perspective of the expert's voice. *Med Educ*. 2011;45:927-38.
- Edelbring S, Broström O, Henriksson P, Vassiliou D, Spaak J, Dahlgren LO, Fors U, Zary N (2012): Integrating virtual patients into courses: follow-up seminars and perceived benefit. *Med Educ*. 2012 46:417-25.
- Ellaway R, Poulton T, Fors U, McGee JB, Albright S (2008): Building a virtual patient commons. *Med Teacher* 2008;30:170–4.

- Eva KW (2005). What every teacher needs to know about clinical reasoning. *Med Educ* 2005;39:98–106.
- Fall LH, Berman NB, Smith S, White CB, Woodhead JC, Olson AL (2005): Multi-institutional development and utilisation of a computer-assisted learning program for the paediatrics clerkship: the CLIPP Project. *Acad Med* 2005;80:847–55.
- Fischer MR, Kopp V, Holzer M, Ruderich F, Juenger J (2005): A modified electronic key feature examination for undergraduate medical students: validation threats and opportunities. *Med Teach* 2005;27:450–5.
- Fischer MR, Hege I, Hörnlein A, Puppe F, Tönshoff B, Huwendiek S. (2008): [Virtual patients in medical education: A comparison of various strategies for curricular integration]. *Z Evid Fortbild Qual Gesundheitsw* 102:648–653.
- Fischer MR (2000): CASUS – An authoring and learning tool supporting diagnostic reasoning. In: Ch. Daetwyler, editor. Use of computers in medical education (Part II). *Zeitschrift für Hochschuldidaktik* 1/2000 pp 87–98.
- Garde S, Bauch M, Haag M, Heid J, Huwendiek S, Ruderich F, Singer R, Leven F. CAMPUS - Computer-Based Training in Medicine as part of a Problemoriented Educational Strategy. *Studies in Learning, Evaluation, Innovation and Development*, 2005;2:10–19. <http://sleid.cqu.edu.au/include/getdoc.php?id=189&article=49&mode=pdf> [Accessed 15 September 2008].
- Grunwald T, Corsbie-Massay C. Guidelines for cognitively efficient multimedia learning tools: educational strategies, cognitive load, and interface design. *Acad Med* 2006;81:213–23.
- Haag M & Huwendiek S: The virtual patient for education and training: a critical review of the literature. *IT-Information Technology* 2010, 52: 281-287
- Higgs, J, Jones M, Loftus S, Christensen N: Clinical reasoning in the health professions. 2008: Elsevier Health Sciences.
- Huang G, Reynolds R, Candler C (2007): Virtual patient simulation at US and Canadian medical schools. *Acad Med*. 2007 May;82(5):446-51.
- Huwendiek S, Reichert F, Bosse HM, de Leng BA, van der Vleuten CP, Haag M, Hoffmann GF, Tönshoff B (2009a): Design principles for virtual patients: a focus group study among students. *Med Educ* 2009; 43:580-8.
- Huwendiek S, deLeng B, Zary N, Fischer MR, Ruiz JG, Ellaway R (2009b): Towards a Typology of Virtual Patients. *Med Teach* 2009; 31: 743-748
- Huwendiek S, Duncker C, Reichert F, De Leng BA, Dolmans D, van der Vleuten CP, Haag M, Hoffmann GF, Tönshoff B (2013): Learner preferences regarding integrating, sequencing and aligning virtual patients with other activities in the undergraduate medical curriculum: A focus group study. *Med Teach* 2013; 35:920-9.
- Huwendiek S, deLeng B, Kononowicz AA, Kunzmann R, Muijtjens A, van der Vleuten CPM, Hofmann GF, Tönshoff B, Dolmans DHJM (2014): Exploring the validity and reliability of a questionnaire for evaluating virtual patient design with a special emphasis on fostering clinical reasoning. *Med Teach*. 2014 Oct 14:1-8. [Epub ahead of print]
- Hrynchak P, Takahashi SG, Nayer M: Key-feature questions for assessment of clinical reasoning: a literature review. *Med Educ*. 2014 Sep;48(9):870-83. doi: 10.1111/medu.12509. Review.
- Issenberg SB, McGaghie WC, Petrusa ER, Lee Gordon D, Scalese RJ 2005: Features and uses of high-fidelity medical simulations that lead to effective learning: A BEME systematic review. *Med Teach* 27:10–28.
- Lawless KBS. Multimedia learning environments: issues of learner control and navigation. *Instr Sci* 1997;25:117–31.
- Mayer R. *Cambridge Handbook of Multimedia Learning*. Cambridge: Cambridge University Press 2005;pp. 31–48.
- McLachlan JC. The relationship between assessment and learning. *Med Educ* 2006;40:716–7.
- Nendaz MR, Bordage G. Promoting diagnostic problem representation. *Med Educ* 2002;36:760–6.
- Norman G (2005). Research in clinical reasoning: past history and current trends. *Med Educ* 2005;39:418–27.
- Norman G, Bordage G, Page G, Keane D (2006): How specific is case specificity? *Med Educ* 2006;40:618–23.
- Page GG, Bordage G (1995): The Medical Council of Canada's Key Features Project: a more valid written examination of clinical decision-making skills. *Acad Med* 1995;70:104–10.

CHAPTER 1

- Posel N, McGee JB, Fleischer DM. Twelve tips to support the development of clinical reasoning skills using virtual patient cases. *Med Teach*. 2015 Sep;37(9):813-8.
- Ritter FE, Nerb J, Lehtinen E, O'Shea TM, editors. 2007. *In order to learn: How the sequence of topics influences learning*. New York: Oxford University Press.
- Sandoval VA, Dale RA, Hendricson WD, Alexander JB (1987): A comparison of four simulation and instructional methods for endodontic review. *J Dent Educ*. 1987;51: 532–538.
- Schmidt HG & Mamede S (2015): How to improve the teaching of clinical reasoning: a narrative review and a proposal. *Med Educ*. 2015 Oct;49(10):961-73. Review.
- Schmidt HG, Norman GR, Boshuizen HPA (1990): A cognitive perspective on medical expertise – theory and implications. *Acad Med* 1990;65 (10):611–21.).
- Stemler L. Educational characteristics of multimedia: a literature review. *J Educ Multimed Hypermed* 1997;6:339–59.
- Thompson GA, Holyoak KJ, Morrison RG, Clark TK (2006): Evaluation of an online analogical patient simulation program. Paper presented at: IEEE Symposium on Computer-Based Medical Systems; 2006; Salt Lake City, Utah
- van Merriënboer JJ, Sweller J. 2010. Cognitive load theory in health professional education: Design principles and strategies. *Med Educ* 44:85–93.
- van der Vleuten CP and Schuwirth LW (2005): Assessing professional competence: from methods to programmes. *Med Educ*. 2005 Mar;39(3):309-17.
- Zary N, Johnson G, Boberg J, Fors UG. Development, implementation and pilot evaluation of a web-based virtual patient case simulation environment–Web-SP. *BMC Med Educ* 2006;6:10.

CHAPTER 2

Establishing guidance for Virtual Patient design

Design principles for virtual patients: a focus group study among students

Sören Huwendiek, Friedrich Reichert, Hans-Martin Bosse, Bas A de Leng,
Cees P M van der Vleuten, Martin Haag, Georg F Hoffmann, Burkhard Tönshoff

Medical Education 2009: 43: 580–588

ABSTRACT:

Purpose: To examine what students perceive as the ideal features of virtual patient design to foster their learning with a special focus on clinical reasoning.

Methods: 104 fifth-year medical students worked through at least 8 virtual patients representing 4 different designs during their paediatric clerkship. The virtual patients were presented in 2 modes and differed in the authenticity of the user interface (with or without graphical support), predominant question type (Long Menu vs. Short Menu questions) and freedom of navigation (relatively free vs. predetermined). Each mode was presented in a rich and a poor version with regard to the use of different media and questions and explanations explicitly directed at clinical reasoning. Five groups of 4-9 randomly selected students (n=27) participated in focus group interviews facilitated by a moderator using a questioning route. The interviews were videotaped, transcribed and analysed. Summary reports were approved by the students.

Results: Ten principles of virtual patient design emerged from the analysis. A virtual patient should be relevant, of an appropriate level of difficulty, highly interactive, offer specific feedback, make optimal use of media, help students focus on relevant learning points, offer recapitulation of key learning points, provide an authentic web-based interface and student tasks, and contain questions and explanations tailored to the clinical reasoning process.

Conclusion: Students perceived the identified design principles as conducive to their learning. Many of these principles are supported by the results of other published studies. Future studies should address the effects of these principles using quantitative controlled designs.

INTRODUCTION

One of the challenges in medical education is to teach students how to apply their knowledge in dealing with clinical problems. Research has shown that students develop clinical reasoning skills by seeing many patients, actively engaging in problem solving and receiving sufficient feedback (Eva 2005, Norman 2005). Virtual patients (VPs) are increasingly being used in medical education (Bryce et al. 1998, Fall et al. 2005, Huang et al. 2007, Leong et al. 2003) as an adjunct to real patients. They have been defined as “an interactive computer simulation of real-life clinical scenarios for the purpose of medical training, education, or assessment”(Ellaway et al. 2008).

The diversity of VPs (Bryce et al. 1998, Fall et al. 2005, Garde et al. 2005, Zary et al. 2006) and the expense of developing them (Huang et al. 2007) underline the importance of careful attention to VP design. Although outside the domain of medical education there is a considerable literature base for a general framework to develop teaching cases, studies evaluating VP designs are few (Bearman & Cesnik 2001) and focus mostly on specific aspects, like communication skills (Bearman et al. 2001). The question as to which features of VP design are most suited to optimising student learning has remained unanswered so far (Ruiz et al. 2008). When designing VPs for learning with emphasis on clinical reasoning the following strategies from the wider literature seem to offer a relevant general framework. These are strategies to: foster clinical reasoning; develop teaching cases; foster learning from simulation and design efficient multimedia tools for cognitive learning. These strategies should also be incorporated in VPs but we know of no studies investigating whether and how this was done and how this affected student learning.

In a recent review of educational strategies to promote clinical reasoning, Bowen (2006) emphasised: asking open-ended questions; single-sentence summaries of patient problems in abstract terms (Nendaz & Bordage 2002); asking for discriminating features of a set of diagnostic hypotheses; early probing for differential diagnoses (Bordage 1999); prioritising diagnoses; encouraging students to compare and contrast diagnostic hypotheses based on real clinical data; typical presentations of different diagnostic hypotheses; and the relative probabilities of different diagnoses (Bowen 2006).

Kim and colleagues (2006) synthesized strategies for developing teaching cases based on 5 core attributes: relevant, realistic, engaging, challenging, and instructional.

Given that VPs are simulations, albeit low fidelity ones, the following recommendations from the BEME meta-analysis on high-fidelity simulation (Issenberg et al. 2005) may also be relevant to VP design: providing feedback, range of difficulty level, multiple learning strategies, capture clinical variation, controlled environment, individualized learning and defined outcomes.

Sound educational strategies and interface design are essential for designing cognitively efficient multimedia learning tools (Grunwald & Corsbie-Massay 2006, Stemler

1997, Mayer 2005, Lawless 1997). While programs for perfecting task skills rely on the tenets of behaviourism, programs to foster knowledge by engaging students in interpreting and reflecting on problems - like our VPs - rely more on constructivism. This is in line with the idea that interactivity in e-learning materials is a prerequisite for promoting active learning and critical thinking (Stemler 1997, Mayer 2005, Lawless 1997, Huang 2005, Cook et al. 2006). In terms of interface design the following guidelines have been proposed: synchronize audio and visual information, eliminate multi-tasking, optimize representations and design an approachable interface, maintain a stable learning environment, and eliminate redundant information, decrease predetermined navigational control as expertise develops, and provide authentic contexts (Grunwald & Corsbie-Massay 2006).

In a meta-analysis of the effects of Internet-based instruction in health professions education, Cook and colleagues (2008) found positive effects compared with no intervention but heterogeneous and generally small effects compared to other instructional methods. Apparently, learning outcomes depend on the quality of the instructional design methods rather than on the medium used (Cronbach 1979). In accordance with what is suggested in Cook's study (Cook et al. 2008, Cook 2005), we compared different e-learning interventions, that is, we examined students' experiences with four different types of VPs designed to promote clinical reasoning.

We conducted a focus group study to explore students' learning experiences with the different VP designs to explore which features are best suited to promote learning, especially of clinical reasoning, and whether and in what way the strategies described in the literature are translatable to VP design.

METHODS

Virtual Patients

CAMPUS is a vocabulary-based VP shell (Garde et al. 2005) enabling easy development and presentation of VPs. It is suitable for studies of VP design, because the authoring system allows integration of different types of questions and media for any item of a VP case. Another advantage is that VPs can be presented in two modes. In this study, four cases were presented to students via the simulative presentation mode (CAMPUS-Classic VPs) characterised by mainly Long Menu questions, a relatively authentic web-based interface (the patient is presented against the background of a drawing of an examination room and students' actions are reflected in graphical features, e.g. regions of the body that are examined appear in colour) and relatively free navigation (Figure 1 A1-4, page 40). In Long Menu questions the learner enters free text after which fitting item(s) from the long list appear from which students can choose (Figure 1 A4, page 40). Lists can contain up to 8000 items. These questions avoid cueing effects by allowing

automated feedback (Schuwirth et al. 1996). Four other cases were presented via the card-based presentation mode (CAMPUS-Card VPs) with mainly Short Menu questions, no graphics-supported web-based interface and a pre-determined work flow (see supporting information, Figure 1 B1-4, page 40). Both presentation modes are categorised as ‘problem-solving’ VPs (Bearman & Cesnik 2001, Bearman et al. 2001). In both modes the learner has to make the same decisions a physician would have to make in real life. Feedback is based on comparison with expert decision and often accompanied with explanations of why a decision was right or wrong.

The VPs also differed in the use of questions and explanations relating to clinical reasoning and the extent of media use (rich versus standard content, Table 1).

Table 1 Virtual patients with rich and standard content

Rich content	Standard content
Questions and explanations relating to clinical reasoning	
In addition to the standard content the following strategies are used to foster clinical reasoning: Open-ended free text questions with an ideal answer as feedback for comparison:	Questions are included that ask students to list or choose the correct differential diagnoses after performing the physical examination and after lab tests/ additional investigations.
- After the initial, brief presentation of the patient and his/her chief complaint the students are asked to state differential diagnoses.	The defining features of the final diagnosis and symptoms of differential diagnoses are named.
- After history and physical examination the students are asked to summarise the case in abstract terms.	No other strategies are used.
- In the course of the case (following history and physical examination, lab tests and other investigations, respectively) students are iteratively asked to state the defining and discriminating features of the differential diagnoses and which diagnosis they consider most likely and why.	
Use of media:	
In addition to media use in the standard condition, the rich content cases also include videos, e.g. of diagnostic procedures, such as lumbar puncture, normal visual results of investigations for comparison with the patient findings and more graphical illustrations.	Only pictures of the patients, visual results of investigations and a few graphical illustrations are included.

Setting

Since 2000, VPs have been an integral part of undergraduate paediatric clinical training at the Faculty of Medicine, Heidelberg University, Heidelberg (Huwendiek et al. 2006). In the course of their four-week paediatric clerkship, 104 fifth-year medical students worked through at least 8 VPs of 4 different designs: 4 CAMPUS-Classic VPs, 2 of rich content and 2 of standard content; and 4 CAMPUS-Card-based VPs, 2 of rich content and 2 of standard content. The students rotated through the four-week clerkship in three consecutive groups of about 35 students.

The procedure of the four weekly VP sessions was as follows: during the first hour 10 students worked through a VP in self-selected groups of 1-3 students. This was followed by a 30-minute discussion of the case by the group of 10 students facilitated by a tutor. The four other VPs were self-study assignments.

All 8 VP cases were relevant to the paediatric course exam. After the first and second student group had completed their rotations, the 8 cases were changed in terms of presentation mode, content richness and tutor guidance to avoid bias due to the topics or the curricular integration of the cases.

Focus groups

After encounters with real patients with the same complaints, students' perceptions of the VPs were explored in focus group interviews, because this approach is appropriate not only for exploring what students are thinking but also why they are thinking that (Krueger & Casey 2000). Focus group discussions in peer groups provide a relatively safe environment in which the imbalance of power between the researcher and the subjects is mitigated. Group discussions foster deeper analysis of views on a topic and may even refine or modify them. In this respect, focus group interviews are superior to individual questionnaires or interviews.

Subjects

We invited all fifth-year medical students who had just completed their paediatric clerkship to participate in the study. Of the 50% (n=52) who were willing to participate, we randomly selected 27 students to attend 1 of 5 focus group interviews thereby ensuring attendance of at least 4 students per session. The participants received a small financial incentive.

Procedure

Five focus group sessions, lasting 2 hours each, were held. Two sessions were held after the first group had completed the paediatric rotation, two sessions immediately followed the rotation of the second group and one session was held after the third group had completed their rotation. The sessions were moderated by one of the authors (xx), who is an experienced facilitator of small group discussions. In order to ensure consistency across groups, a questioning route was used to guide the interviews (Krueger & Casey 2000) (see supporting information, Table 4, page 41). During the sessions the assistant moderator (FR) took notes and provided support in videotaping the session. The videotapes were transcribed literally. Summary reports of the discussions were sent to the students for approval. Because no significant new ideas emerged during the fifth focus group session, no further sessions were organised.

Analysis

The transcripts were analysed by two of the authors (SH and FR). They first read the transcripts identifying relevant themes. Because not all of the emerging themes reflected one question of the questioning route, the transcripts were thematically sorted and summarised. The analysers (SH and FR) discussed and, if required, adjusted the categories. The students received a summary report of their group session, including illustrative quotations with students' initials in parentheses. Whenever possible the strength of an opinion and group support in terms of frequency were also indicated. All summary reports were approved.

Results

We present principles of VP design based on the students' views of the preferred VP design for learning with emphasis on clinical reasoning. Characteristic citations from the focus group discussions are presented in tables 3 and 4 to illustrate each design principle. We found no major differences in the five focus groups in terms of results generated.

Principles of virtual patient design

1. Relevance

According to the students VPs should be relevant in terms of the incidence of the patient problem and/or the seriousness of the consequences of inappropriate management. They preferred VPs with problems they were unlikely to encounter during clinical training (Table 2, page 33).

2. Appropriate level of difficulty

The level of complexity should be appropriate for students' knowledge levels. Guidance, e.g. through expert comments or predetermined workflow, was more important at first than later when students had more experience (Table 2).

3. Interactivity

In general, VPs should be highly interactive and include a variety of question formats. The students valued Long Menu questions because these are the most effective in engaging their attention. This type of interactive question was also perceived as most realistic, best suited to revealing gaps in students' knowledge and thus best able to motivate students to study (Table 2).

4. Specific feedback

Students wanted to receive feedback on all their decisions to help them understand what an expert would do. The importance of decisions should be specified, for example

whether an investigation is 'obligatory' or 'optional', to help students estimate the value of different investigations in different situations. Additionally, quantitative formative feedback on overall performance in different parts of the work up (medical history, physical examination, differential diagnosis, lab tests, etc.) was considered desirable to help students identify areas where improvement was needed (Table 2).

5. Appropriate media use

Different media (pictures, video, audio) should be used to make a VP as realistic as possible, especially for the physical examination and additional investigations. Presenting a picture or a video of the patient at the start of a case and during history taking was also thought to enhance authenticity. When media were used to present test results (e.g. chest X-ray), 'normal' findings should be presented alongside them to help students recognise discriminating features. Media, such as animations and graphical illustrations, were preferred whenever they were considered superior to words in explaining things (Table 2).

6. Focus on relevant learning points

The web-based interface should help students focus on relevant learning points. This could be achieved by striving for clarity, avoiding distractions and ensuring ease of use of the VP. Text should be kept as brief and clear as possible and access to additional information should be facilitated by hyperlinks or context sensitive search functions. The main learning points should be highlighted, for instance by presenting them in bold or colour (Table 3).

7. Recapitulation of key learning points

At the end of a case, questions should be asked to evaluate what students have learned and to recapitulate important points. The VP should conclude with a summary including key take home messages and students should be able to print out any relevant information (Table 3).

8. Authenticity of the web-based interface

The web-based interface should be authentic to make students feel like the doctor in charge thereby engaging and motivating them. The physical examination should be visualised by media supported simulation to help students proceed in a structured manner (e.g. from head to toe) (Table 3).

9. Authenticity of student tasks

A VP should require students to make all the clinical decisions a physician would make, such as which history questions to ask, which parts of the physical examination to perform, which laboratory tests and additional investigations to order, whether the patient can be managed as an outpatient or should be admitted to the intensive care unit or

hospital ward and all decisions relating to differential diagnosis, management and treatment. Long Menu questions with a realistic number of options and open-ended questions were perceived as most realistic and most motivating (see also design principle 3) (Table 3).

10. Questions and explanations to enhance clinical reasoning

Students should be actively involved in the clinical reasoning process in the same way an expert would be. This can be achieved by asking students to generate differential diagnoses early in the case, by eliciting discriminating features of a set of diagnostic hypotheses through open-ended free text questions and by providing appropriate feedback. Early synthesis should be encouraged by having students summarise relevant findings and providing them feedback on this. Emphasis should be on evaluating students' understanding of the crucial steps in patient management and on appropriate feedback. The students indicated that explicit questioning and guiding of the clinical reasoning process as well as design principles 3, 4 and 9 made them feel well prepared for dealing with real patients to an extent no other type of teaching had achieved before (Table 3).

Table 2 Illustrative citations from the focus group discussions about principles 1-5 of virtual patient design
The numbers in parentheses indicate the focus group and the student

1. Relevance

- *"I found it helpful that the mandatory cases were typical and important cases and not very rare or complicated ones. Because I had worked through a virtual patient, I knew, perhaps for the first time, what to ask and do when I saw a real patient with similar symptoms in an outpatient setting."* (2,5)
- *"One of the things I liked is that via the virtual patients I could 'take care of' a patient with meningitis, which would not have been possible in reality."* (1,2)
- *"It would be great to have cases of the ten most common diseases for each medical discipline"* (3,5)

2. Appropriate level of difficulty

- *"At first I needed more guidance from the program on how to proceed through the case and found this very helpful. As I became more proficient I needed less guidance and enjoyed knowing how to proceed on my own."* (1,4)
 - *"I found it hard to solve a CAMPUS-Classic case in the first week of the paediatric clerkship. I would rather have started with a CAMPUS-Card case, which is easier because of the short menu questions and more guidance"* (3,1)
 - *"At first, you should have the rather easy cases, later the more difficult ones."* (4/3)
-

3. Interactivity

- *"I think a variety of question types, such as multiple choice, long menu or free text questions are helpful in making the case interactive, making you think, keeping you focused. At first multiple choice question are suitable, because your knowledge base is rather small. Later long menu questions should be used to challenge students' more active knowledge."* (2,5)
- *"I think the long menu questions are especially suitable for testing if you understand the patient problem and can actively apply your knowledge. They showed me the gaps in my knowledge and motivated me the most to study."* (3,5)
- *"The long menu questions often caused "aha" effects. For example if you had to give five differential diagnoses, you very quickly recognised the gaps in your knowledge, which was not the case with the short menu questions."* (4,2)

4. Specific feedback

- *"It is very helpful to receive feedback from an expert on every decision to learn what is right and what is wrong and why it is so."* (1,3)
- *"It is very helpful when feedback identifies crucial aspects you must never forget and less important aspects".* (4,1)
- *"It was good to get feedback on your overall performance in the different parts of the patient work up, to know in which areas you are doing well and where improvement is required."* (4,4)

5. Appropriate media use

- *"The initial picture of the patient, the illustration of physical findings with pictures and video sequences and pictures of the results of lab tests and other investigations helped to make the case realistic and get the feeling that there is a real patient in front of you."* (4,2)
- *"It is very helpful when a finding in which you can see something is also illustrated by a picture and then you have to describe it and get feedback. That is much better for learning than to have the findings directly reported in words. Furthermore I learned more from graphics illustrating complex relationships than from texts describing complicated relationships."* (1,2)
- *"It was very helpful to also see normal findings for comparison, e.g. a normal chest X-ray or normal lung sounds. The more often you can make this comparison, the better you get at recognising the important differences."* (4,3)

Table 3 Characteristic citations from the focus group discussions about principles 6-10 of virtual patient design
Numbers in parentheses indicate the focus group and the student

6. Focus on relevant learning points

- *"The web-based interface should be such that it is self-evident and everything is easy to find and use and not distracting"* (1,3)
- *"I think texts should focus on the most relevant issues and be clearly written and possibly supported by tables and graphics. The main issues should be marked by printing the text in bold or different colours."* (5,7)
- *"To make texts more clear, they should focus on the essential issues only and link to further details via hyperlinks or by searching in a textbook or database."* (4,2)

7. Recapitulation of key learning points

- *"I found the many multiple choice and free text questions at the end of a case very helpful to recapitulate important aspects and check whether I really understood things."* (3,3)
- *"I found the summary with the take home messages at the end of the case very helpful. It should be possible to print out this page and all the other relevant comments, figures etc."* (5,2)
- *"I got the most out of questions like 'to what 5 things do you pay the most attention in this or that aspect?' And then I had to write that down myself".* (3/2)

8. Authenticity of the design of the web-based interface

- *"[In CAMPUS-Classic VPs with a more authentic web-based interface] I felt that I first made a decision and then carried it out. It felt like being the doctor in charge, sitting behind my desk, and I had to treat this patient. This motivated me."* (3,5)
- *"I found it useful to have the picture of the child in front of me and then perform a structured physical examination. It is more realistic and authentic that way and helps to get an own standard in performing a physical."* (4,4)
- *"[The more authentic web-based interface] with its visualisation of the treatment room and the physical examination supported the feeling of being the doctor in charge much more than if mainly textual information is provided."* (4/3)

9. Authenticity of student tasks

- *"It was great to manage the patient from admission to discharge and to make all the decisions, which we had done much too rarely before this, but which we should practise far more often in other clinical specialities as well."* (4,2)
- *"I found the simulative system [CAMPUS-Classic VPs] with long menu questions very realistic, because as in real life you have to know things and you do not get a limited number of options to choose from. This motivated me greatly to study and resulted in higher learning gains."* (2,3)
- *"In real life you rarely have the opportunity to follow a patient from admission to discharge, let alone make all the decisions, including what to ask, investigate and how to manage the patient. To do this and get feedback was extremely helpful"* (4/3)

10. Questions and explanations tailored to the clinical reasoning process

- *"It helped my learning when questions, expert comments and feedback focused on clinically relevant themes, such as the differential diagnostic approach and patient management."* (2,2)
- *"I found the questions that focused on clinical management and differential diagnosis very helpful. In combination with the feedback they helped me to develop a picture of what are relevant differential diagnoses in such a patient and why and which are not. This kind of focus on differential diagnosis I missed in my previous studies. When I met a real patient with a similar symptom in an ambulatory setting I knew for the first time exactly what to do and in what order"* (3,3)
- *"I found it very helpful that right from the beginning the program actively involved you in a thinking process like that of a doctor. For example, at the start, we were asked to give our differential diagnosis and then again and again after ruling out the differential diagnoses step by step and stating the correct diagnosis."* (1,2)
- *"It helped enormously when the program asked you to state what supports the differential diagnoses and what does not and why. So you become aware of which information or investigations are relevant for your decisions which you may not have been aware of if the program had not asked you about it explicitly. In this respect, I found the free text questions ideal."* (3,3)
- *"I found it helpful that you were asked to summarise the case in abstract words. It helped you focus your thinking on differential diagnoses in a certain direction and directed your further decisions."* (5,7)

DISCUSSION AND CONCLUSION

We explored students' perceptions of VPs to identify features that make VPs suitable for learning, especially related to clinical reasoning. The students discussed VP design after exposure to at least 8 VPs varying in presentation mode and in the amount of explicit attention to clinical reasoning and media usage. According to the students, learning is facilitated when VPs are relevant, of appropriate difficulty, interactive, provide specific feedback, make appropriate use of different media, help them focus on

relevant learning points, foster recapitulation of key learning points, are authentic in terms of the web-based interface and student tasks and ask questions and offer explanations that enhance clinical reasoning.

Some of the criteria mentioned by the students are in line with design criteria for multimedia learning tools described in the literature, such as interactivity (principle 2, (Stemler 1997, Huang 2005)). However, to our knowledge, no studies have compared the effects of Short and Long Menu questions on learning with multimedia learning tools. Our study shows that students favour Long Menu questions, because they find these questions challenging, suitable for identifying gaps in their prior knowledge and motivating.

According to Grunwald and Corsbie-Massay (2006), web-based interfaces address two conflicting goals: that of entertaining and engaging the learner, and that of remaining free of potentially distracting elements to limit cognitive load and ensure that any changes are lasting and cumulative. This is exactly what the fifth year medical students told us about the web-based interface design when they asked for authenticity on the one hand (principle 8) but also for focus on relevant learning points achieved through absence of redundant information, ease of use, and highlighting of important issues with links to more detailed information (principle 6). Although our study focused on features of VP design that foster clinical reasoning in particular, some features mentioned by the students also reflect general principles of effective case design (principles 1, 2, 4), as described by Kim et al. (2006).

According to our students, it is not only the web-based interface that should be authentic (principle 8) and supported by authentic media use (principle 5). Authenticity is even more crucial for the tasks students are given to perform (principle 9). The students want to be prompted in a realistic way to make all the required clinical decisions (principle 9) and receive appropriate guidance and feedback (principles 2 and 4). This view is also voiced by Herrington (2006), who in his article on authentic e-learning concludes that “cognitive realism” is rather more important for learning than photo-realistic design of simulations.

In the introduction section, we referred to Bowen’s (2006) recent review of strategies to enhance students’ clinical reasoning skills. The students found these strategies (principle 10), together with principles 3, 4 and 9, especially helpful in fostering clinical reasoning and preparing them for encounters with real patients. To our knowledge, there are no reports in the literature of the implementation of such strategies in virtual patients and their impact on student learning. Our results also demonstrate that VPs designed in accordance with the principles proposed in this study are judged by students as supporting their learning and offering excellent preparation for clinical reasoning in real patient encounters.

This study is limited because it was confined to the lowest Kirkpatrick level. Naturally, the effects of the VPs on student learning also need to be studied by controlled experimentation using outcome variables higher up the Kirkpatrick levels. Another limita-

tion is that focus group interviews is a qualitative method which is susceptible to bias, e.g. by the influence of the moderator. We tried to counteract this by having an experienced moderator actively promote input from all participants, stressing that different experiences were equally valuable and asking the students to write down some thoughts on a topic before it was discussed in the group. Because the abundance of VP systems and design options precluded inclusion of all possible VP designs, it cannot be ruled out that some relevant aspects may have been overlooked. Another limitation is that the generalizability of the results is limited because they were obtained from clerkship-level undergraduate medical students only.

Contributors: SH conceived, designed and implemented the study, analysed the data and wrote the first draft of the paper. FR assisted in the implementation of the study and analysis of data.

H-MB, BT, GFH, MH, CPMvdV and BAdL contributed to the design of the study. All authors participated in the writing of the paper.

Acknowledgements: the authors thank Professor F J Leven, M Bauch, J Heid and F Hess of the Centre for Virtual Patients for their co-operation in the development of the two-player versions, and M L B Gorsira for critically reading and correcting the English manuscript.

Funding: this study was supported by a grant from the Medical Faculty of Heidelberg University, Heidelberg, Germany.

Conflicts of interest: none.

Ethical approval: not required. We confirm that participation was voluntary, the participants cannot be identified from the material presented and no plausible harm to participating individuals can arise from the study.

REFERENCES

- Bearman M, Cesnik B (2001): Comparing student attitudes to different models of the same virtual patient. *Medinfo*. 2001;10:1004-1008.
- Bearman M, Cesnik B, Liddell M (2001): Random comparison of 'virtual patient' models in the context of teaching clinical communication skills. *Med Educ*. 2001;35:824-832.
- Bordage G. Why did I miss the diagnosis? Some cognitive explanations and educational implications. *Acad Med*. 1999;74:S138-143.
- Bowen JL (2006): Educational strategies to promote clinical diagnostic reasoning. *The New England journal of medicine*. 2006;355:2217-2225.
- Bryce D, King N, Graebner C, Myers J (1998). Evaluation of a diagnostic reasoning program (DxR): Exploring Student perceptions and addressing faculty concerns. *Journal of interactive media in education* [cited 1998, 2008 Jun 1];1:[1-35 screens]. Available from: <http://www.jime.open.ac.uk/98/1/bryce-98-1.pdf>.
- Cook DA (2005): The research we still are not doing: an agenda for the study of computer-based learning. *Acad Med*. 2005;80:541-548.
- Cook DA, Thompson WG, Thomas KG, Thomas MR, Pankratz VS (2006): Impact of self-assessment questions and learning styles in Web-based learning: a randomized, controlled, crossover trial. *Acad Med*. 2006;81:231-238.
- Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM (2008): Internet-based learning in the health professions: a meta-analysis. *Jama*. 2008;300:1181-1196.
- Cronbach LS SR (1997): *Aptitudes and Instructional methods*. New York Irvington Publishers; 1979.
- Ellaway R, Poulton T, Fors U, McGee JB, Albright S (2008): Building a virtual patient commons. *Medical teacher*. 2008;30:170-174.
- Eva KW (2005): What every teacher needs to know about clinical reasoning. *Med Educ*. 2005;39:98-106.
- Fall LH, Berman NB, Smith S, White CB, Woodhead JC, Olson AL (2005): Multi-institutional development and utilization of a computer-assisted learning program for the pediatrics clerkship: the CLIPP Project. *Acad Med*. 2005;80:847-855.
- Garde S, Bauch M, Haag M, Heid J, Huwendiek S, Ruderich F, Singer R, Leven F (2005): CAMPUS - Computer-Based Training in Medicine as part of a Problemoriented Educational Strategy. *Studies in Learning, Evaluation, Innovation and Development*, 2005;2:10-19. <http://sleid.cqu.edu.au/include/getdoc.php?id=189&article=49&mode=pdf>. [Accessed 15 September 2008].
- Grunwald T, Corsbie-Massay C (2006): Guidelines for Cognitively Efficient Multimedia Learning Tools: Educational Strategies, Cognitive Load, and Interface Design. *Acad Med*. 2006;81:213-223.
- Herrington J (2006): Authentic e-learning in higher education: Design principles for authentic learning environments and tasks. *World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*, Honolulu, HI, USA, 2006.
- Huang C (2005): Designing high-quality interactive multimedia learning modules. *Computerized medical imaging and graphics*. 2005;29:223-233.
- Huang G, Reynolds R, Candler C (2007): Virtual patient simulation at US and Canadian medical schools. *Acad Med*. 2007;82:446-451.
- Huwendiek S, Köpf S, Höcker B, Heid J, Bauch M, Bosse HM, Haag M, Leven FJ, Hoffmann GF, Tönshoff B (2006): Fünf Jahre Erfahrung mit dem curricularen Einsatz des fall- und webbasierten Lernsystems "CAMPUS-Pädiatrie" an der Medizinischen Fakultät Heidelberg. *GMS Z Med Ausbild*. 2006;23: Doc 10. <http://www.egms.de/pdf/journals/zma/2006>
- Issenberg SB, McGaghie WC, Petrusa ER, Lee Gordon D, Scalese RJ (2005): Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Medical teacher*. 2005;27:10-28.
- Kim S, Phillips WR, Pinsky L, Brock D, Phillips K, Keary J (2006): A conceptual framework for developing teaching cases: a review and synthesis of the literature across disciplines. *Med Educ*. 2006;40:867-876.

- Krueger RA, Casey MA (2000): Focus groups: a practical guide for applied research. 3 ed. London Sage Publications, Inc; 2000.
- Lawless KBS (1997): Multimedia learning environments: Issues of learner control and navigation. *Instructional Science*. 1997;25:117-131.
- Leong SL, Baldwin CD, Adelman AM (2003): Integrating Web-based computer cases into a required clerkship: development and evaluation. *Acad Med*. 2003;78:295-301.
- Mayer R (2005): *Cambridge Handbook of Multimedia Learning*. Cambridge Cambridge University Press.; 2005.
- Nendaz MR, Bordage G. Promoting diagnostic problem representation. *Med Educ*. 2002;36:760-766.
- Norman G (2005): Research in clinical reasoning: past history and current trends. *Med Educ*. 2005;39:418-427.
- Ruiz J G ZN, Huang G, Huwendiek S, Leng de B, McGee J, Vollmar H, Waldmann U-M, Flacker J, Soriano R, Smith S, Issenberg B. BEME systematic review (2008): Virtual patients. *Proceedings of the Conference of the Association for Medical Education in Europe*; Prague. 2008.
- Schuwirth LW, van der Vleuten CP, Stoffers HE, Peperkamp AG (1996): Computerized long-menu questions as an alternative to open-ended questions in computerized assessment. *Med Educ*. 1996;30:50-55.
- Stemler L (1997): Educational Characteristics of Multimedia: A Literature Review. *Journal of Educational Multimedia and Hypermedia*. 1997;6:339-359.
- Zary N, Johnson G, Boberg J, Fors UG (2006): Development, implementation and pilot evaluation of a Web-based Virtual Patient Case Simulation environment--Web-SP. *BMC medical education*. 2006;6:10.

SUPPORTING INFORMATION:



Legend Figure 1:

- A1-A4: Simulative presentation mode (CAMPUS-Classic)
- B1-B4: Card-based presentation mode (CAMPUS-Card).
- A1 vs. B1: Relatively authentic web-based interface (CAMPUS-Classic) *versus* non-authentic web-based interface (CAMPUS-Card). A click on one of the icons “mother with child”, “stethoscope”, “X-ray on wall” in A1 takes a student to the related part of the case work up, such as history taking, physical examination and additional investigations etc.
- A2 vs. B2: Same case and content of the two presentation modes.
- A3 vs. B3: Physical examination is visualised in the simulative presentation mode (CAMPUS-Classic) on a mannequin *vs.* no visualisation (CAMPUS-Card); the pictures of the patient or of remarkable findings are the same in both presentation modes.
- A4 vs. B4: Long Menu *vs.* Short Menu questions.

Table 4 The sequence of questions (questioning route) used by the moderators in guiding the focus group discussions

We would like you to think back to when you were working through the different virtual patients.

Question 1 (30 minutes)

- Which aspects of the two virtual patient player modes fostered your learning (especially clinical reasoning) the most and why. Try to illustrate this with some examples.

Question 2 (15 minutes)

- Which aspects of the two virtual patient player modes did not support your learning and why. Try to illustrate this with some examples.

Question 3 (30 minutes)

- Which aspects of the content of the virtual patients promoted your learning (especially clinical reasoning) the most and why. Try to illustrate this with some examples.

Question 4 (10 minutes)

- Which aspects of the content of the virtual patients did not support your learning and why. Try to illustrate this with some examples.

Question 5 (35 minutes)

- Suppose there are unlimited resources for developing virtual patients, how should the ‘ideal’ virtual patient be designed to maximise your learning especially in clinical reasoning and prepare you best for your future professional practice, i.e. dealing with real patients. Try also to think of aspects you have not experienced yourself. Try to be as specific as possible.
-

CHAPTER 3

Typing Virtual Patient Design

Towards a typology of virtual patients

Sören Huwendiek, Bas A. deLeng, Nabil Zary, Martin Fischer,
Jorge G. Ruiz, Rachel Ellaway

Medical Teacher 2009: 31: 743-748

ABSTRACT

Introduction: Although on-screen “virtual patients (VPs)” have been around for decades it is only now that they are entering the mainstream, and as such they are new to most of the medical education community. There is significant variety in the form, function, and efficacy of different VPs and there is, therefore, a growing need to clarify and distinguish between them. This article seeks to clarify VP concepts and approaches using a typology of VP designs.

Methods: The authors developed a VP design typology based on the literature, a review of existing VP systems, and their personal experience with VPs. This draft framework was refined using a Delphi study involving experts in the field, and was then validated by applying it in the description of different VP designs.

Results: Nineteen factors were synthesized around four categories: general (title, description, language, identifier, provenance, and typical study time); educational (educational level, educational modes, coverage, and objectives); instructional design (path type, user modality, media use, narrative use, interactivity use, and feedback use); technical (originating system, format, integration, and dependence).

Conclusion: This empirically derived VP design typology provides a common reference point for all those wishing to report on or study VPs.

BACKGROUND

A number of medical educators have been enthusiastically embracing the use of virtual patients for teaching and learning activities (Huang et al, 2007). However, all too often educators and developers use the term virtual patient ambiguously. Language shapes and delimits our understanding, and having a robust and common set of terms helps to identify and understand differences and similarities between different instances of phenomena. A key step to better understanding and reporting therefore is the development of well-defined and unambiguous language and concepts to describe (and therefore provisionally model) the phenomena under observation. This step should be considered in the context of problems with comprehensive and common vocabularies for medical education as a whole (Haig et al, 2004) and with virtual patients in particular (Cook and Triola, 2009). The framework presented in this paper is intended to improve precision and clarity in research, development and practical applications of virtual patients.

VIRTUAL PATIENTS

A virtual patient has been defined as "an interactive computer simulation of real-life clinical scenarios for the purpose of healthcare and medical training, education or assessment" (Ellaway, Candler et al., 2006). This definition is intentionally broad and inclusive and therefore not without problems, not least because it covers so many different kinds of designs and applications. The development and use of virtual patients is not a new phenomenon, there have been variations on a theme of computer simulations of patient encounters since the 1970s when computers first became available in classroom settings and there have been many tools and materials developed since then.

It should be noted that although virtual patients can be built manually using standard web and multimedia tools it is more usual (at least so far) for them to be built inside dedicated systems that both streamline the authoring process and provide the facilities to run the resulting virtual patients. As with any technology, these systems intrinsically reflect the norms, values and concepts held by their designers (Scarborough and Corbett, 1992) and the kinds of virtual patients they support in turn reflect the designs and affordances of their originating systems.

Returning to the definition of what constitutes a virtual patient we are immediately faced with a problem:

- Interventions and designs that fall within the description of virtual patients are described using different terms. For instance the following terms (with variations) have all been used to describe designs that are what we would consider to be virtual patients: 'case-based learning systems' (Aha, 1991, Garde et al. 2005), 'computer-aided simulations of the clinical encounter' (Harless et al, 1971: Melnick, 1990), 'in-

teractive patients' (Gerritsma and Smal, 1988: Hayes and Lehman, 1996), 'computerized clinical patient problems' (Pickell et al, 1986), 'patient simulation by computer' (Verbeek, 1987: Clauser et al, 2002), 'patient simulations' (Bergin and Fors, 2003: Finkelstein et al, 1991) and finally 'virtual patients' (Zary et al, 2006: Begg et al, 2007).

- Interventions that may be called virtual patients but are quite different from the working definition, including: 'simulated patients' or "standardized patients" (actors playing the role of patients, for instance in developing communication skills or as part of an OSCE station), physical simulation artefacts (ranging from whole body models such as mannequins, to individual body parts such as task trainers), physiological models (computerized algorithms representing physiological processes) and real patients reflected in the data held about them in health records and other databases.

Clearly, clarifying what interventions are similar and which are different is a major part of informed development and a move to better and more targeted use of this class of educational interventions. A well-defined typology is necessary to clarify the structure of virtual patients, discriminate between the different virtual patient types and respective instructional approaches, avoid the confusion which exists when a single term covers a host of sometimes incompatible instructional and assessment approaches, facilitate the search process of virtual patients by authors, teachers and learners, for purposes of virtual patient design and for research and evaluation. By being able to differentiate virtual patient designs and approaches, it becomes possible to more objectively and meaningfully ascertain the advantages and disadvantages of each and where they might most appropriately be used.

Towards a Typology

The work presented here arose from the EU eViP Project (<http://www.virtualpatients.eu>), that involved the exchange of virtual patient materials between eight institutions using four different systems and spanning six countries. The technical interoperability was tackled using the MedBiquitous virtual patient specification (http://medbiquitous.org/working_groups/virtual_patient) as a transport mechanism between systems. However, a key part of this project was to develop clearer mutual understanding and appreciation of the different approaches taken by the different partners, most of who already had significant track records in virtual patient development.

All of the partners had extensive experience in developing and using virtual patients but in different ways. A key requirement was therefore to build mutual understanding of the different approaches reflected in the systems used:

- CASUS (<http://www.casus.net>): developed at Ludwig-Maximilians-University Munich and the spin-off company Instruct AG, is a multilingual authoring and runtime shell

based on a client-server architecture. Virtual patients developed with the authoring component can be presented in a card-based mode for learning with eight different answer-types and a pre-determined workflow (path type: linear string of pearls). A mapping-tool for supporting hypothetical-deductive processes is also included. The assessment mode on a dedicated server supports secure browsing and encrypted logs of user performance. For more on this system see Fischer (2000).

- CAMPUS (<http://www.campusvirtualpatients.com>): developed at the University of Heidelberg, is a multilingual vocabulary-based VP shell. VPs developed with the authoring component can be presented in three modes: (1) The simulative mode (CAMPUS classic) with mainly long menu questions, an authentic user interface and relatively free navigation (path type: linear string of pearls); (2) The card-based mode (CAMPUS card) with mainly short menu questions and a pre-determined work flow (path type: linear string of pearls); and (3) the assessment mode with a network fault tolerant client-server-architecture and certainty of law using automatically generated video files. For more on this system see Garde et al. (2005).
- OpenLabyrinth (<http://sourceforge.net/projects/openlabyrinth/>): originally developed at the University of Edinburgh and subsequently by a number of partners as an open source educational gaming engine it uses branching paths and a range of game-like features such as counters, rules, conditionals and dynamic feedback to support a wide range of different scenario and activity designs. It should be noted that other than supporting the MedBiquitous data standard there is nothing medicine-specific in the OpenLabyrinth platform. For more on this system see Begg et al (2007).
- Web-SP (<http://www.web-sp.org/>): developed at the Karolinska Institutet in Stockholm is a multi-lingual framework for creating and running virtual patients aimed at all health science disciplines. It supports linear string of pearls paths composed of a combination of the modules available. For more on this system see Zary et al. (2006)

Other systems and approaches considered include those developed at McGill Molson (<http://www.afmc.ca/chec-cesc/cases-e.php>), Tufts (<http://tusk.tufts.edu/about/caseplayer>), Pittsburgh (<http://zone.medschool.pitt.edu/sites/LET/ardp.aspx#VP>), NYU (<http://wise-md.med.nyu.edu/Init.action>), IVIMEDS (<http://www.ivimeds.org/Ivimeds/pagepublic.do?id=7>), and Stanford (<http://timhowgego.com/stanford-virtual-worlds-research.html>).

Other reference points included:

- The IEEE learning object metadata (LOM) specification (<http://ltsc.ieee.org/wg12/>) designed for the classification of educational materials. The LOM has useful typological patterns such as those for describing contributions, semantic density and versioning.
- Meller's typology of simulation (1997) identifies components within a simulation regarding what is being simulated (the patient and/or disease process, procedures,

tests, equipment, physician or paraprofessional and the professor or expert practitioner) and in what way (passive, active, interactive).

- Kim and colleagues (2006) identified a set of key factors for developing teaching cases, including; relevance (level of learner, goals and objectives, setting of case narrative) realism (authenticity, distractors, gradual disclosure), engagement (rich content, multiple voices and perspectives, outcomes dependent on learners' actions), challenge (difficulty, atypicality, variation in structure and sequencing), and instructional (scaffolding, assessment, tracking, feedback, resources and aids).

The development of the typology was executed as follows:

- The authors initially developed a 'straw man' proposal for a virtual patient design typology based on the reference frameworks, a review of existing virtual patient systems, and their personal experience with virtual patients. This first iteration was split into design and implementation sections with 23 factors identified in the former and 9 factors in the latter.
- The next step involved the first draft framework being refined by repeatedly asking for comments, changes and ratings of the different factors in a Delphi like study within the author team to create a second framework. Contributors rated the importance of each domain as well as each of the corresponding factors on a 5-point numerical scale. Raters were instructed to indicate those items whose meaning or importance they did not understand or agree with. The second framework combined the design and implementation aspects into four headings (basic characteristics, design, educational, technical) with 27 factors across these headings.
- Finally, a third framework was derived and validated by applying the second framework in the description of a subset of different exemplar virtual patients and involving further virtual patient experts. When all expert members had completed the third round of factor and domain rating, mean ratings were calculated and further adjustments were made to reflect the consensus gained.

All six authors completed the three rounds of the Delphi consensus building process. Iterative discussions were held between the authors during which the overall structure of the typology was discussed and developed. Other inputs involved individual review of other virtual patient systems encountered at other institutions or from commercial providers and a background review of the literature gathered as part of a BEME review into virtual patients.

RESULTS

The typology derived is expressed as a framework that can be used in both as a descriptive model and as a way of classifying. This resulting framework involves 19 key factors grouped under four main categories (see table 1).

Table 1: Virtual patient typology framework

Category	Factor	Description
General	Title	Name given to the virtual patient activity.
	Description	Free-text
	Language	Standard language indicator including version – eg EN-GB indicates British English and EN-US is US English.
	Identifier	Unique ID or address
	Provenance	Author and other contributor information.
	Typical study time	Time in minutes for typical learner to complete activity.
Educational	Educational level	Target learner level, for instance year 1 undergraduate, 3 rd year resident etc
	Educational modes	Teaching, learning and assessment, and formative or summative aspects.
	Coverage	Topic area(s) covered.
	Objectives and outcomes	Objectives of the activity and any outcomes they address.
Instructional Design	Path type	Linear string of pearls, branching
	User modality	Number of users involved in an activity, roles they take
	Media & resources	Use of images, audio, video, animations etc.
	Narrative use and patient focus	Balance between presenting data (reading a history, getting test results) and telling a story (engaging in conversation, character, motive).
		In what way are patients involved in the activity design – are they the main focus (such as diagnosing and treating a patient) or a vehicle for other issues such as team working, basic science or professionalism.
	Interactivity use	Kinds of questions, tasks and challenges faced. Overall number of cognitive interactions.
	Feedback use	Kinds of feedback and whether during an activity or at the end (or both)
Technical	Originating system	Which tool or system was used to create the activity?
	Format	What technical format is used at the user end – typically text/HTML for web and application/EXE for disc-based media.
	Integration and dependence	What other tools, systems or other contextual factors does the activity depend upon to run properly?

The framework was validated against the authors' own systems and cases (table 2).

Table 2: High level review of virtual patients from four different systems using the typology framework.

	CASUS	CAMPUS classic	OpenLabyrinth	Web-SP
Title	Sneezing pet dealer with foot pain	Infant with fever	Sarah Jane	Tom P Miller
Description	Elderly female pet dealer with acute pain in her right foot in the absence of an adequate trauma.	Infant with fever comes to a paediatric emergency room. Learner is the doctor in charge for this patient and has to take the same decision a doctor would need to make in real life (e.g. choosing history questions, physical exams, lab tests, DD, therapies, etc.).	Learner diagnoses and treats a sick baby and deals with her distraught mother.	Patient with symptoms of pneumonia
Language	German	German	EN-GB	EN-US
Identifier	evip:vp:1000170	evip:vp: 1000263	Multiple versions	evip:vp:1000114
Provenance	Eversmann, Ruf, Fischer, LMU Munich	Huwendiek, Höcker, Seidel, University of Heidelberg	Jonathan Round, St Georges, University of London	Virtual Patient Lab, Karolinska Institutet, Sweden
Typical study time	20 minutes	45 minutes	15 minutes	25 minutes
Educational level	Undergraduate – 3 rd year	Undergraduate – final years	Undergraduate – final years	Undergraduate – clinical years
Educational modes	Learning and formative assessment	Learning and formative assessment	Formative assessment	Learning and assessment
Coverage	Internal Medicine	Paediatrics	Paediatrics	Internal medicine
Objectives and outcomes	Learners will learn the different reasons for pathologic fractures. They will also learn about risk factors, preventive measures, and diagnostic and therapeutic strategies with respect to those fractures with a focus on osteoporosis.	Learners will learn how to gather relevant information, diagnose, treat and manage an infant with fever. At the end of the activity learners will be able to: (1) Describe how to establish a diagnosis in an infant with fever without focus; (2) Describe how to rule out relevant differential diagnoses; (3) Describe how to manage such an infant.	Learners will rehearse skills diagnosing and managing a sick baby along with other environmental issues.	Learners will learn to diagnose a patient with pneumonia and provide the proper treatment and management strategies
Path type	Linear string of pearls	Linear string of pearls	Branching	Linear string of pearls

	CASUS	CAMPUS classic	OpenLabyrinth	Web-SP
User modality	Single user plays the role of an MD in the emergency room of a hospital, follow up in GP practice.	Single user plays the role of the paediatrician in charge, first in the emergency room then on ward	Single user plays the role of a GP then an SHO	Single user plays the role of an MD at a university hospital
Media & resources	Text, static graphics and images.	Text, static graphics, videos	Purely text-based	Text, images, video and sound
Narrative use and patient focus	Told from the patient's perspective	Told from patient's and mother's voices active.	Told from second-person perspective, patient and mother's voices active.	Told from the patient's perspective
Interactivity use	Clickable options (questions, exams, labs) and text input (diagnosis, patient management) Overall number of cognitive interactions: 10	Long menu questions Multiple choice questions Free text questions Overall number of cognitive interactions: 133	Clickable text options for the different decision points in the scenario.	Clickable options (questions, exams, labs) and text input (diagnosis, patient management) Overall number of cognitive interactions: 57
Feedback use	Feedback is given to each decision by comparison with the expert decision. Often reasons why a decision is right or wrong are provided. Feedback is provided immediately and statistically summarized at the end of the activity.	Feedback is given to each decision by comparison with the expert decision. Often reasons why a decision is right or wrong are provided. Feedback is provided immediately and statistically summarized at the end of the activity.	Formative as part of case narrative – simple concluding points when endpoint is reached, no score or grading.	Delayed feedback – both formative (neutral and individualized) and summative (built-in scoring/grading module)
Originating system	CASUS, - see Fischer (2000)	CAMPUS classic - see Garde et al. (2005)	Excel/OpenLabyrinth – see Round (2006).	Web-SP Built-in authoring tool
Format	DHTML frontend, servlet and JSP backend.	Interactive Java-Applet inside Web-Browser	Text/HTML	Text/HTML
Integration and dependence	Requires CASUS backend to run	Java Plug-in respectively Java Runtime Environment (JRE), CAMPUS backend, otherwise independent	Requires OpenLabyrinth to run, otherwise independent	Requires Web-SP to run, otherwise independent

DISCUSSION

In this empirically-derived virtual patient design typology we synthesized 19 factors around 4 categories: general (title, description, language, identifier, provenance, typical study time); educational (educational level, educational modes, coverage, objectives); instructional design (path type, user modality, media use, narrative use, interactivity use, feedback use); technical (originating system, format, integration and dependence).

In developing this typology we have identified a number of issues, which we will address in the following.

- The instructional design of virtual patients depends much on the technical affordances of the systems used to create them. We can envisage a two-step model that describes the functionality within the host system followed by specific characteristics of the virtual patient(s) within that system. It is also possible that if and as the key affordances required of virtual patient systems become clearer that there will be greater convergence between systems through convergent adoption as core functionality. For now, there is significant (although now better understood) divergence between system designs, which in turn limits functional separation between the instructional design and their authoring and delivery context. It should be noted that this has also been a key dynamic in the development of the MedBiquitous interoperability model for virtual patients with developers implementing key features in the specification into their own platforms.
- Activities such as PBL cases and virtual patients have much of their value linked to progressive disclosure and discovery. Declaring the factors “coverage” and “objectives and outcomes” of the category “educational” in total to learners in advance could significantly diminish this value. Public descriptions of any educational resource need to be circumspect as to what is disclosed in advance so as to retain their emergent value.
- Generally, and despite their different approaches to virtual patient development, there was a good level of consensus between the authors in developing the framework. For instance, some discussion was required to clarify relatively unfamiliar or ill-defined areas such as narrative, patient focus or user modality than to do with opposing needs or perspectives. One area requiring further investigation is that of the instructional design in terms of the path type of virtual patients. Two broad categories were identified (see figure 1):
 - Linear string of pearls: there is a predefined sequence of master steps through a case, each one can contain many options. There is a single endpoint.
 - Branching: the learner can select from any number of branching alternatives through a case possible with different end points

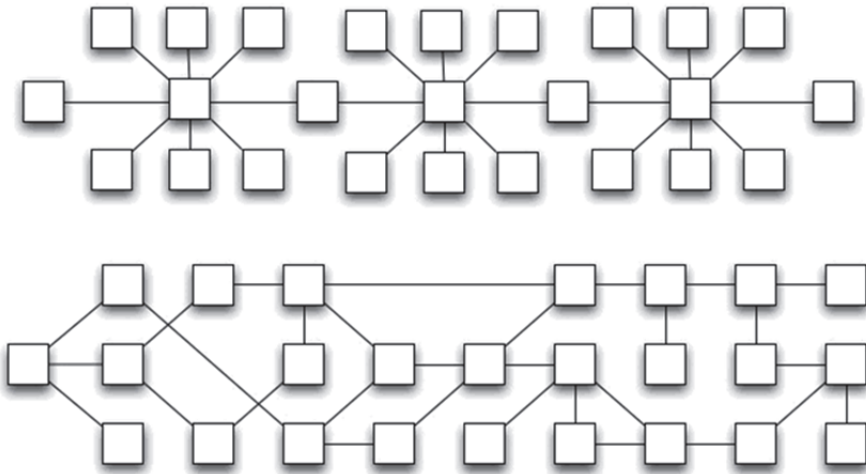


Figure 1: Two path type variations of virtual patients (category: instructional design). Boxes represent user actions, choices or screens and the lines between them the available options to move between them. At top is the linear string of pearls model and at the bottom the branching model.

A number of the constituent factors of the proposed typology have existing empirical and theoretical bases in strategies for teaching case design (Kim and colleagues 2006), for instance appropriate educational level, media and interactivity use, and the use of feedback. New factors include factors such as narrative and user modalities and creating a comprehensive model that combines educational and technical factors, in particular the path type of virtual patients

The framework is intended to foster enhanced discourse among virtual patient developers and users, particularly with respect to understanding similarities and differences between different systems and instances. For instance, a number of the authors are also involved in a BEME review of virtual patients; the work reported here is presented to be an essential precursor study to the main review. The framework is also intended to help clarify key concepts and support greater rigour in research into the development and use of virtual patients. In particular, a common semantic framework can support the much needed experimental investigation and identification of causal links between virtual patient design factors and their utility and efficacy as raised by Cook and Triola (2009). We encourage greater critical awareness and clarity in understanding and differentiating between these new educational interventions, in design, application and evaluation.

A limitation of this study is that although a number of different designs were considered in creating this framework it is not exhaustive in sampling from all such designs worldwide. It is also recognized that virtual patient development is not static and that new approaches, technologies and educational models are still being developed, for

instance involving the use of virtual worlds such as Second Life. As a result, it is expected that this framework will continue to be refined and improved. The next step will therefore be to implement the framework with a larger group of virtual patient developers and experts and investigate their reactions to the framework.

CONCLUSIONS

Despite the increasing use of virtual patients worldwide there has so far been no overarching study that synthesizes the many designs and approaches to create a common framework and vocabulary of virtual patient design. The authors, by developing an empirically-derived virtual patient design typology, expressed as an integrated framework, have provided a common reference point for all those wishing to report on or study virtual patients.

Practice Points

- Virtual Patients are becoming part of mainstream medical education
- There are no suitable frameworks or vocabularies at present for describing virtual patients in a way that clarifies their similarities and differences
- This paper reports a virtual patient framework based on 19 factors across four categories
- The proposed virtual patient design typology can help to clarify virtual patient concepts and approaches and establish more objective and meaningful ways of reporting on or evaluating them

Notes on contributors

Sören Huwendiek, MD, MME (Bern) is Chair of the Centre for Virtual Patients and e-learning commissioner at Heidelberg Medical School. His research focuses on the design and curricular integration of virtual patients. He is a lead participant in the eViP Project.

Bas de Leng works as Assistant Professor at the department of Educational research and Educational development of the Faculty of Health, Medicine and Life Sciences of Maastricht University. His research focuses on blended learning in PBL with special interest in Computer Supported Collaborative Learning and Technology-Enhanced Learning. He is a lead participant in the eViP Project.

Nabil Zary, PhD, is Associate Senior Lecturer in Medical Simulation with focus on Virtual Patient research, Dept of Learning, Informatics, Management and Ethics, Karolinska Institutet, Stockholm, Sweden. He is a lead participant in the eViP Project.

Martin R. Fischer, MD, MME (Bern) is Professor at the Institute for Teaching and Educational Research in Health Sciences at the Medical Faculty of the Private University Witten/Herdecke. His research focuses on computer supported case-based learning and assessment. He is a lead participant in the eViP Project.

Jorge G. Ruiz, MD, FACP is Associate Professor of Clinical Medicine, Director Geriatrics Fellowship Program Division of Gerontology and Geriatric Medicine University of Miami Miller School of Medicine GRECC Associate Director for Education/Evaluation VA Medical Center GRECC. He is the lead on a BEME systematic review on virtual patients (in preparation).

Rachel Ellaway, BSc, PhD is Assistant Dean and Associate Professor of Education Informatics at the Northern Ontario School of Medicine and Visiting Professor at St Georges University of London. She is also co-chair of the MedBiquitous Virtual Patient Working Group and Chair of the AFMC Informatics Resource Group. She acts in a consultative role to the eViP project.

REFERENCES

- Aha, D.W. (1991) Case-based learning algorithms. In Proceedings of the DARPA Case-Based Reasoning Workshop (pp147-158). Washington, D.C.: Morgan Kaufmann <http://citeseer.ist.psu.edu/aha91casebased.html> [accessed 30 December 2008].
- Begg, M., Ellaway, R., Dewhurst, D., Mcleod, H. (2007) "Transforming Professional Healthcare Narratives into Structured Game-Informed-Learning Activities." *Innovate* 3(6). Retrieved May 11, 2009, from <http://www.innovateonline.info/index.php?view=article&id=419&action=article>
- Bergin, R. and Fors, U. (2003) Interactive Simulation of Patients – an advanced tool for student-activated learning in medicine & healthcare. *Computers and Education*, 40/4, pp361 – 376.
- Clauser BE, Margolis MJ, Swanson DB. (2002) An examination of the contribution of computer based case simulations to the USMLE step 3 examination. *Acad Med*. 77(10):S80-82.
- Cohen, L., Manion, L. et al. (2000) *Research Methods in Education*. UK, Routledge.
- Cook, D.A., Triola, M.M. (2009) Virtual patients: a critical literature review and proposed next steps. *Medical Education* 43(4) pp303-311
- Ellaway, R., Candler, C., Greene, P., Smothers, V. (2006) *An Architectural Model for MedBiquitous Virtual Patients*. Baltimore, MD, MedBiquitous.
- Fischer, M.R. (2000). CASUS - An Authoring and Learning Tool Supporting Diagnostic Reasoning. In: *Use of Computers in Medical Education (Part II)*. Ch. Daetwyler (Ed.), *Zeitschrift für Hochschuldidaktik* 1/2000 pp87-98.
- Finkelstein, M.W., Johnson, L. A., and Lilly, G.E. (1991) A computer management system for patient simulations. *Computer Methods and Programs in Biomedicine*, 34, pp257–261.
- Garde S., Bauch M., Haag M., Heid J., Huwendiek S., Ruderich, F., Singer R., Leven F. J. (2005) CAMPUS - Computer-Based Training in Medicine as part of a Problem-oriented Educational Strategy. *Studies in Learning, Evaluation, Innovation and Development*, pp10–19. Retrieved May 11, 2009, from <http://sleid.cqu.edu.au/viewissue.php?id=6>
- Gerritsma, J.G.M. and Smal, J.A. (1988) An interactive patient simulation for the study of medical decision-making. *Medical Education* 22(2): pp118-123.
- Haig, A., Ellaway, R., Dozier, M. (2004) "METRO the creation of a taxonomy for medical education." *Health Information & Libraries J* 21(4): 211-219.
- Harless, W.G., Drennon, G.G., Marxer, J.J., Root, J.A., and Miller, G.E. (1971) CASE: a computer-aided simulation of the clinical encounter. *J. Med. Educ.*, May, 46(5), pp443–448.
- Hayes, K.A., and Lehman, C.U. (1996) The interactive patient: a multimedia interactive educational tool on the World Wide Web. *MD Computing*, 13(4), pp330–334.
- Huang G, Reynolds R, Candler C. (2007) Virtual patient simulation at US and Canadian medical schools. *Acad Med*. 82(5): pp446-51.
- Kim S, Phillips WR, Pinsky L, Brock D, Phillips K, Keary J (2006). A conceptual framework for developing teaching cases: a review and synthesis of the literature across disciplines. *Medical Education* 40(9): pp867–876.
- Meller, G. (1997) A typology of simulators for medical education. *J Digit Imaging*. 10(3) pp194-6.
- Melnick D. (1990) Computer-Based Clinical Simulation: State of the Art. *Eval Health Prof*. 13:104-120.
- Pickell, G.C., Medal, D., Mann, W.S. and Stabler, R.J. (1986). Computerizing Clinical Patient Problems: an evolving tool for medical education. *Medical Education* 20(3): pp201-203.
- Round, J. (2006). Creating your own virtual patients. 01 (The newsletter of the Higher Education Academy Subject Centre for Medicine, Dentistry and Veterinary Medicine) http://www.medev.ac.uk/external_files/pdfs/01_newsletter/0113_lo_res.pdf.
- Scarborough, H. and J. M. Corbett (1992). *Technology and organization: power, meaning and design* UK, Routledge.
- Verbeek, H.A. (1987). Self-instruction through patient simulation by computer. *Medical Education* 21(1): pp10-14.

Zary, N., Johnson, G., Boberg, J., and Fors, U. (2006) Development, implementation and pilot evaluation of a Web-based Virtual Patient Case Simulation environment – Web-SP. BMC Medical Education, <http://www.biomedcentral.com/1472-6920/6/10>.

CHAPTER 4

Evaluating Virtual Patient design

Exploring the validity and reliability of a questionnaire for evaluating virtual patient design with a special emphasis on fostering clinical reasoning

Sören Huwendiek, Bas A. deLeng, Andrzej A. Kononowicz, Romy Kunzmann,
Arno Muijtjens, Cees P.M. van der Vleuten, Georg Friedrich Hofmann,
Burkhard Tönshoff, Diana H.J.M. Dolmans

Medical Teacher 2014; Oct 14:1-8. [Epub ahead of print]

ABSTRACT

Background: Virtual patients (VPs) are increasingly used to train clinical reasoning. So far, no validated evaluation instruments for VP design are available.

Aims: We examined the validity of an instrument for assessing the perception of VP design by learners.

Methods: Three sources of validity evidence were examined: (i) Content was examined based on theory of clinical reasoning and an international VP expert team. (ii) The response process was explored in think-aloud pilot studies with medical students and in content analyses of free text questions accompanying each item of the instrument. (iii) Internal structure was assessed by exploratory factor analysis (EFA) and inter-rater reliability by generalizability analysis.

Results: Content analysis was reasonably supported by the theoretical foundation and the VP expert team. The think-aloud studies and analysis of free text comments supported the validity of the instrument. In the EFA, using 2547 student evaluations of a total of 78 VPs, a three-factor model showed a reasonable fit with the data. At least 200 student responses are needed to obtain a reliable evaluation of a VP on all three factors.

Conclusion: The instrument has the potential to provide valid information about VP design, provided that many responses per VP are available.

INTRODUCTION

Virtual Patients (VP) are increasingly being used in medical education (Fall et al. 2005, Huang et al. 2007, Poulton & Balasubramaniam 2011, Berman et al. 2011, Cook et al. 2010, Cook & Triola 2009). They have been defined as ‘interactive computer simulations of real-life clinical scenarios for the purpose of medical training, education, or assessment’ (Ellaway et al. 2008), and seem to be especially suited to fostering clinical reasoning (Cook & Triola 2009). VP design is essential for the educational success of VPs (Huwendiek et al. 2009b, Edelbring et al. 2012). The inherent costs of developing VPs (Huang et al. 2007) underline the importance of paying careful attention to VP design. Together with the increasing use of VPs both nationally (Fall et al. 2005, Berman et al. 2008 & 2011) and internationally (Balasubramaniam et al. 2009, Poulton & Balasubramaniam 2011), within the scope of the “electronic Virtual Patients” (eVIP) project (www.virtualpatients.eu/, Poulton et al. 2007), these arguments led us to the realization that a short and standardized evaluation tool of VP design is a prerequisite for achieving an further expansion of VP use in medical education.

To the best of our knowledge, to date, there is no existing theory-guided and validated evaluation instrument for assessing VP design that would allow improved judgments and feedback in the process of developing VPs. For the development of such an evaluation instrument, a theoretical framework is required (Cook & Beckman 2006), but no such framework has yet been published.

As a conceptual basis for the development of an evaluation instrument for assessing VP design, we found the following model and strategies to be useful:

Gruppen and Frohna described a model of clinical reasoning that integrates the key features of several theoretical frameworks (Gruppen & Frohna 2002). In summary, this model encompasses six elements:

- 1) Patient situation/characteristics: This describes the patient, with all of his/her symptoms and characteristics.
- 2) Context: This describes the circumstances in which patients present their problems and/or questions, because clinical reasoning cannot occur in a vacuum (Gruppen and Frohna 2002).
- 3) Information gathering: This describes an active inquiry and collection of relevant information to judge a patient’s condition.
- 4) Problem representation: Through his/her prior knowledge, the physician experiences the patient in his own way (gets a mental picture of the patient) and discriminates between important and rather less important information. The representation of the clinical problem changes throughout the clinical reasoning process.
- 5) Evaluation: This means the continuous comparison of the problem representation and the real situation and characteristics of the patient. This allows the physician to decide whether further information is required to establish the correct diagnosis;

- 6) Prior knowledge: The prior knowledge of the physician is a very important component of the clinical reasoning process. The organization of knowledge has been shown to be highly relevant for expertise development (Gruppen & Frohna 2002).

The above-mentioned elements of clinical reasoning are usually passed through several times in a cyclical manner before a decision or diagnosis is finally reached.

In a recent review of educational strategies to promote clinical reasoning, Bowen emphasised techniques including the following (Bowen 2006): asking open-ended questions; asking for single-sentence summaries of patient problems in abstract terms; asking for discriminating features of a set of diagnostic hypotheses; comparing and contrasting diagnostic hypotheses based on real clinical data; and demonstrating typical presentations of different diagnostic hypotheses and the relative probabilities of different diagnoses.

Kim et al. (2006) synthesised strategies for developing teaching cases based on five core attributes and concluded that teaching cases need to be relevant, realistic, engaging, challenging and instructional.

Huwendiek and colleagues (Huwendiek et al. 2009b) described the following principles for designing VPs to foster clinical reasoning: A VP should be relevant, have an appropriate level of difficulty, be highly interactive, offer specific feedback, make optimal use of media, help students to focus on relevant learning points, offer recapitulation of key learning points, provide an authentic web-based interface and student tasks, and contain questions and explanations tailored to the clinical reasoning process.

Taking these theoretical notions and strategies into account, an instrument to evaluate VP design was developed within the eVIP project (Poulton et al. 2007, Huwendiek & de Leng 2010), which aimed to develop and share high-quality VPs among European countries. The aim of the present study is to find support for the validity of this instrument.

To assess the validity of an instrument, a variety of different sources of evidence can be distinguished (American Educational Research Association and American Psychological Association 1999, Downing 2003, Cook & Beckmann 2006): (i) Content: do instrument items completely represent the construct? (ii) Response process: the relationship between the intended construct and the thought processes of subjects or observers; (iii) Internal structure: acceptable reliability and factor structure; (iv) Relations to other variables: correlation with scores from another instrument assessing the same construct; (v) Consequences: do scores really make a difference?

In this study, we examine the first three of these sources of validity evidence to determine the reliability and validity of an instrument which is filled in by medical students and which aims at assessing the perception of VPs designs with a special focus on fostering clinical reasoning.

METHODS

Context

The VP design instrument was developed within the international eVIP expert consortium (Poulton et al. 2007, Huwendiek & de Leng 2010, www.virtualpatients.eu) to allow the evaluation of VP design with an emphasis on clinical reasoning using a standardised instrument. The final instrument should allow an evaluation of VP design during regular courses at the involved medical schools. One prerequisite was that the instrument should be easy and quick to fill in. Therefore, the decision was made to use a small number of rather broad questions accompanied by free text questions rather than a large number of detailed and specific questions. This should allow for a broad use of the instrument and enable a good response rate to be obtained. A further prerequisite was that the instrument should ultimately be available in the partners' native languages (English, Dutch, German, Polish, Romanian and Swedish).

Content

As there is no single inclusive model for the evaluation of VPs available, as described in the introduction, we used the existing literature to inform the theoretical framework of this instrument. A preliminary instrument was developed prospectively by two of the authors (SH & BdL), taking into account the following literature: The model of clinical reasoning by Gruppen & Frohna (2002) formed the main basis for three theoretical factors of the instrument (see below). As instructional aspects are not part of Gruppen and Frohna's (2002) model, we added items regarding instructional methods to our model, which can be used within VPs to foster learning especially of clinical reasoning. These were informed by Bowen (2006), Kim et al. (2006) and Huwendiek et al. (2009a). Furthermore, a fourth factor regarding the "learning effect of consultation" was added in order to ascertain whether students learn from VPs. The English draft versions were sent to all VP experts of the eVIP consortium (n=9) for feedback. Through several rounds of written and verbal feedback, the instrument was further developed and modified (particularly in terms of rewording several items), until consensus on the factors and items of the instrument was reached.

The developed items were clustered around the following factors (the respective elements of the factors are indicated in brackets):

- Factor 1: Authenticity of the patient encounter and the consultation (context, patient characteristics (Gruppen & Frohna 2002)).
- Factor 2: Cognitive strategies in the consultation (problem representation, evaluation, information gathering (Gruppen & Frohna 2002), instructional methods (Bowen 2006, Huwendiek et al. 2009b)).

Factor 3: Coaching during consultation (prior knowledge (Gruppen & Frohna 2002), instructional methods (Kim et al. 2006, Huwendiek et al. 2009b).

Factor 4: Learning effect of consultation (learning effect).

Students were asked to respond to each of the 11 items in the instrument based on a 5-point Likert scale ranging from 1 – ‘strongly disagree’ to 5 – ‘strongly agree’. Students were also asked to rate the overall learning experience on the same scale. Each question was accompanied by a free text question, asking students why they answered the question as they did. Furthermore, the following three open-ended questions were added at the end of the questionnaire: (i) special strengths of the case, (ii) special weaknesses of the case, (iii) any additional comments.

To guarantee that the content of the final consensus-based evaluation instrument was comprehensible for students, think-aloud protocols with students were applied (for details, see “response process”, below). After any necessary refinements, the questionnaire was then translated into the five languages (Dutch, German, Swedish, Polish and Romanian) in which the instrument had to be applied. To guarantee that these local versions of the questionnaire still reflected the original constructs, think-aloud protocols were again performed with local students and the instruments were modified accordingly. Finally, the instruments were back-translated into English to check that the translations corresponded to the original English instrument. The final instruments were included as an electronic survey at the end of the VPs of the eVIP project, which allowed for easy data collection and analysis. The final versions in six languages are also available via the eVIP homepage (<http://www.virtualpatients.eu/resources/evaluation-tool-kit/translated-versions/>). See the final English version in the appendix.

Response process

The response process was investigated using two methods: a pilot study with students using a think-aloud technique and a content analysis of free text answers accompanying each closed question.

Think-aloud pilot studies

Students’ think-aloud studies were performed at two time points: first with the final consensus-based English-language instrument with native English-speaking medical students (n=6), which was then refined accordingly, and second with the translated instruments. In both instances, students were asked to think aloud while filling in the instrument, to discuss the relevance and wording of the items, and to identify factors that affected their answers. These group discussions lasted for about 90 min. The instruments were refined accordingly to achieve the intended understanding of the questions. These discussions resulted in the rewording of several items in both instances.

Content analysis of free text answers accompanying each closed question

In addition, each item in the instrument was accompanied by a free text question asking students why they answered a question as they did. The German free text answers were analysed using content analysis (Mayring 2010). Two researchers (RK & SH) discussed the results and modified the coding accordingly. Themes and subthemes were established in an iterative process, in which the coded themes informed subsequent discussions, and the discussions in turn informed the coding process. The researchers discussed and, if considered appropriate, adjusted the themes until consensus was reached. The percentage of answers coded to themes corresponding to the theoretical foundation (elements) of each of the factors was calculated as an indicator of whether students had understood the questions in the intended manner.

Internal structure

The multilingual instrument was used by 2,547 medical students in three European countries (the Netherlands (Maastricht), Germany (Heidelberg), Poland (Krakow)) for the judgement of a total of 78 VPs.

First, mean scores and standard deviations were computed at the item level and at the factor level. Computing mean scores is appropriate, as research has shown that the Likert response format produces interval data at the scale level (Carifio & Perla 2008).

A scale construction procedure was carried out to validate the instrument. Scales were obtained by repeated exploratory factor analysis (EFA), reliability analysis, and content analysis. The factor structure was investigated for the data at the VP level because it concerns inter-item relations at the level of the rated subject. Items were removed if item loadings were ambiguous (difference less than 0.2 for loadings on different factors) and/or item loadings were too low (less than 0.4) and/or if the reliability (Cronbach's alpha) of the scale improved when the item was removed. The coefficient alpha was estimated for each factor to determine its internal consistency. A coefficient of 0.70 or higher was considered as acceptable. This step of the analysis results in subsets of items (scales). In the next steps, the mean score among the items of a scale (scale score) is used as a variable to be analysed.

Inter-rater reliability

Generalizability studies were conducted to investigate the inter-rater reliability of the scale scores resulting from the previous step, as well as for the single item global score. The analyses were conducted for the data at the rating level (student response) for the 78 VPs (2,547 responses of Dutch, German and Polish medical students). The generalizability studies show how many student responses are needed per VP for a reliable indication of the VP's quality (Crick & Brennan 1983). A students-nested-within-groups design was used, with VPs as the object of measurement. This design allows variance component estimation of two sources: (i) differences between VPs (object of measure-

ment) and (ii) differences between students nested within VP judgements (Shavelson & Webb 1991). Reliability indices (generalizability (G)) coefficient and standard error of measurement (SEM) are reported as a function of the number of students rating a VP.

RESULTS

Content

As described in the methods section, a theoretical framework was used to develop the questionnaire. Furthermore, an international VP expert team was involved to secure the content validity of the instrument. However, it has to be taken into account, that both above mentioned arguments are rather weak indications for the content validity of the instrument.

RESPONSE PROCESS

Think-aloud study

The think-aloud study led to the rewording of several items of the instrument. Addition or removal of items was not deemed to be necessary. The think-aloud studies ensured the intended understanding of the questionnaire as described in the methods section. All students agreed that the evaluation instrument is relevant and useful for evaluating VPs. Students supported the idea of administering the questionnaire online as a link at the end of a VP.

Content analysis of free text comments accompanying each item of the instrument

Percentages of free text comments corresponding to the theoretical foundation of each of the factors of the instrument were as follows: Factor 1 (Authenticity): 89% (373/418), Factor 2 (Cognitive strategies): 74% (239/323), Factor 3 (Coaching): 90% (497/549), Factor 4 (Learning effect): 90% (71/79). These high percentages of free text comments corresponding to the theoretical foundation of each of the factors of the instrument indicate that students understood the items of the questionnaire in the intended manner. Examples of the free text comments can be found in Table 1.

Table 1: Response process, content analysis of free text comments, examples of free text comments. The themes and subthemes to which the free texts were coded are written in brackets.

Factor 1: Authenticity of patient encounter and the consultation

“clinically oriented case with the same possibilities as in a hospital for investigations” (theme: context; subtheme: realistic diagnostic and therapeutic options/decisions of the case)

“realistic patient case” (theme: patient characteristics; subtheme: realism of the case itself)

Factor 2: Cognitive strategies in the consultation

“because you always had to think about whether new results of investigations would change your working diagnosis” (theme: evaluation; subtheme: judgement of the overall plausibility of the working diagnosis)

“you had to ask for the right information like history questions, lab tests etc. to get the right information for evaluating the diagnosis” (theme: information gathering; subtheme “information gathering for the evaluation of the working diagnosis”)

Factor 3: Coaching during consultation

“the good clarification in the feedback was very helpful for me in how to come to the right diagnosis in the patient” (theme: instructional methods; subtheme: good clarifications in feedback)

“the questions were appropriate and allowed for active thinking about the differential diagnosis” (theme: prior knowledge; subtheme: appropriateness and helpfulness of questions)

Factor 4: Learning effect of consultation

“I felt more experienced after doing this case, because I repeated relevant aspects” (theme: learning effect of the consultation; subtheme: learning effect because of repetition of relevant aspects)

“After working on this case I have the differential diagnostic approach in my head and will always think about the five important differential diagnoses in an infant with fever without focus” (theme: learning effect of consultation; subtheme: learning effect because of focus on differential diagnostic approach)

Internal structure

Seventy-eight VPs were evaluated by 2547 students in the Netherlands, Germany and Poland. The exploratory factor analysis showed that the original model – based on all four factors (see Appendix) – did not fit the data. After eliminating five items, including both items of the factor “learning success of the consultation”, a three-factor model for the remaining six items was shown to represent the structure in the data reasonably well. Final factor analysis for these six items resulted in a Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO) = 0.61 (mediocre level), and an explained variance of 84%. Alpha coefficients per factor were 0.74 – 0.82, indicating a high inter-item reliability per scale (see table 2). The mean scores and SDs for the three factors and for the global score at the VP level are reported in Table 2. Item mean scores and standard deviations (SDs) for the data at the rater (student) level (N=2547) are shown in Table 3 for the items of the final three-factor model.

Table 2: Number of items and VP-level (N=78) statistics (minimum and maximum score, mean score (scale 1-5), SD, and coefficient alpha) for the scale scores of the final three-factor model, and the global score.

	Item(s)	N items	N VP	Min.	Max.	Mean	SD	Alpha
F 1: Authenticity	1,2	2	78	3.05	4.17	3.55	0.20	.81
F 2: Cognitive strategies	4,6	2	78	3.30	4.33	3.82	0.21	.74
F 3: Coaching	8,9	2	78	3.21	4.50	3.87	0.21	.82
Global Score	12	1	78	3.43	4.69	4.02	0.26	-

Table 3: Mean and SD per item in the final three-factor model for the returned evaluations (N=2547).

Factors and items	Mean(1-5)	SD
Factor 1: Authenticity of patient encounter and the consultation		
While working on this case, I felt I had to make the same decisions a doctor would make in real life.	3.62	0.84
While working on this case, I felt as if I were the doctor caring for this patient.	3.39	0.89
Factor 2: Cognitive strategies in the consultation		
While working through this case, I was actively engaged in revising my initial image of the patient's problem as new information became available.	3.74	0.82
While working through this case, I was actively engaged in thinking about which findings supported or refuted each diagnosis in my differential diagnosis.	3.80	0.77
Factor 3: Coaching during consultation		
The questions I was asked while working through this case were helpful in enhancing my diagnostic reasoning in this case.	3.82	0.81
The feedback I received was helpful in enhancing my diagnostic reasoning in this case.	3.89	0.78
Global Score		
Overall, working through this case was a worthwhile learning experience.	3.98	0.76

The results of the generalizability studies demonstrate that the variance associated with the VPs for the global score is 5.86%, as reported in Table 4. This percentage is the true variance or the variance of interest. The variance associated with VPs varies per factor between 1.26% and 3.73%. The estimated variance components were used to estimate reliability indices (G-coefficients). Table 4 provides, per factor, the G-coefficient and the standard error of measurement (SEM) as a function of the number of student responses (N). The SEM can be used to estimate confidence intervals for individual scores.

The SEM should be lower than 0.25 in order to achieve a 95% confidence interval (estimated score $\pm 2 \times \text{SEM}$), including only one of the scale points 1, 2, 3, 4, 5. To obtain a reliable G-coefficient of at least 0.70 or higher, at least 40 students' responses are needed for the Global Score, 75 student responses are needed for factors 2 and 3, and at least 200 student responses are required to obtain reliable results for factor 1 (Table 4).

Table 4: Estimated variance components for the variance associated with VPs and raters (students) are given (in percentages in brackets) and the generalizability coefficient (G-coefficient) and standard error of measurement (SEM), as a function of the number of student ratings (N) for the 3 factors (scale 1–5) and the Global Score

Factor	Estimated variance components		N	G-coefficient	SEM
	inter-VP variance	rater variance within VP			
F1: Authenticity	0.007 (1.26%)	0.575 (98.74%)	50	0.39	0.11
			75	0.49	0.09
			100	0.56	0.08
			200	0.72	0.05
F2: Professional approach	0.014 (3.03%)	0.455 (96.97%)	50	0.61	0.10
			75	0.70	0.08
			100	0.76	0.07
F3: Coaching	0.018 (3.73%)	0.470 (96.27%)	50	0.66	0.10
			75	0.74	0.08
			100	0.79	0.07
Global Score	0.034 (5.86%)	0.542 (94.14%)	40	0.71	0.12
			50	0.76	0.10
			75	0.82	0.09
			100	0.86	0.07

Note: The global score was a separate item and is not the mean score of all of the items.

DISCUSSION

This article focuses on the validity and reliability of an instrument to assess VP design with a special emphasis on fostering clinical reasoning by students. We sought evidence concerning the first three sources of evidence from the model of standards proposed (American Educational Research Association and American Psychological Association 1999, Downing 2003, Cook & Beckmann 2006).

As the instrument was developed involving VP experts and based on a theoretical framework of clinical reasoning, we have some indication for content validity. The validity evidence relating to the response process in the think-aloud pilot studies led to the rewording of items and ensured the intended understanding of the questionnaire by the end users. Furthermore, the comparison of the themes in the free text argumentation of the students in the final instrument (why they rated a question in a positive or negative manner) with the elements of the underlying theoretical framework showed that high percentages of answers referred to elements of the respective factors of the theoretical framework. This indicates that students understood the items in the intended manner, providing additional response process validity.

The results of the exploratory factor analysis indicated that a three-factor model fits the data reasonably well (Table 2). In our study, the factor “learning success of consultation”, as the 4th factor of the original instrument, did not fit the model. A reason for this

might be that questions concerning one's own learning success are particularly rater-dependent.

The findings of the generalizability studies indicated that at least 40 students' responses are needed for the global score, 75 student responses are needed for factors 2 and 3, and at least 200 student responses are required to obtain reliable results for factor 1 (Table 4). It is disappointing that at the factor level, about 200 student responses are needed to obtain reliable data for all factors. The reason for this high number lies in the low inter-VP variance on the authenticity of the patient encounter (see also Table 4). This might be because the authenticity of a VP (factor 1) is somehow relative and students in different places, with different cultures, and with different experiences with real patients were included in the analysis. Another interpretation why high numbers of student ratings are needed might be that the instrument is not capable of detecting enough true VP variance due to the use of a limited number of rather broad questions, which was done deliberately to allow for a good acceptance of the instrument. This might have been the wrong choice and more specific items might lead to better discrimination. Further research is in order here.

However, when using the instrument as an electronic survey after each VP, large numbers of ratings should not be a problem in universities with high numbers of students. Thus, based on the exploratory factor analysis and the generalizability studies, it can be concluded that the instrument appears to reveal reasonably valid factor scores, provided that there are enough evaluations per VP available.

The finding that the factor scores of the instrument are reasonably valid and reliable, provided that enough evaluations are available, implies that the data collected can be used to measure student perception of the value of a VP's design in teaching clinical reasoning. This is important, as the design of VPs has been shown to be relevant for students' learning (Huwendiek et al. 2009b, Cook et al. 2010). Thus, the information obtained using the instrument can create awareness and an increased understanding of student perception of VP design. This awareness and evaluation can encourage teachers to improve the design of VPs.

The strength of this study is that, to our knowledge, it is the first to validate an instrument for the assessment of the perception of VP design, developed based on a theoretical construct and with a focus on clinical reasoning, which is usually regarded as the main aim of learning with VPs (Cook & Triola 2009). Furthermore, the instrument was developed by international VP experts and validated in an international sample of students using three sources of validity evidence. Finally, the instrument, with only seven questions (six items, one global score), is convenient to use by students and can be feasibly applied in educational practice. First experiences in using the evaluation results generated by this instrument for improving VPs are promising.

A limitation of the instrument is that many evaluations are needed to obtain reliable and valid judgements for all factors. In view of the increasing use of the same VPs by very large numbers of students both nationally (Fall et al. 2005, Berman et al. 2008 &

2011) and internationally (Balasubramaniam et al. 2009, Poulton et al. 2007, Poulton & Balasubramaniam 2011), the required numbers of evaluations should mostly become achievable. Additional research is needed to assess the use of this instrument in terms of providing insight into the influence of multicultural and multilingual settings on VP design perception. Further validation of the instrument with a larger number of students is also recommended.

CONCLUSION

This study provided information regarding three sources of evidence for the validity of the developed instrument. The three-factor VP design evaluation instrument appears to be an instrument with a theoretical foundation, which has reasonable validity evidence for the evaluation of the perception of VP design by students with a special focus on clinical reasoning, provided that enough evaluations per VP are available. The validated instrument in this study can be used to provide teachers with feedback about the perception of VP design in large student cohorts, e.g. by an electronic survey at the end of a VP, allowing for quick data collection and evaluation, and enabling timely changes in VP design.

Practice points

- The design of VPs is crucial for student learning. In view of the increasing use of VPs, it is essential to evaluate their design.
- The VP design evaluation instrument in this study appeared to reveal reasonable validity provided that enough evaluations are available.
- The instrument can be used to provide teachers with information about student perception of a VP's design in teaching clinical reasoning

Declaration of interest:

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the article.

Acknowledgements

We thank all of the other former eVIP partners and VP experts (<http://www.virtualpatients.eu/>) who are not co-authors of this paper, namely Chara Balasubramaniam, David A Davies, Martin Fischer, Valentin Muntean, Terry Poulton, Uno Fors and Nabil Zary for their support in developing this instrument and Sarah Mannion de Hernandez for the

English language correction of the manuscript. Furthermore, we would like to express our gratitude to all of the students who filled in the questionnaires

NOTES ON CONTRIBUTORS:

Sören Huwendiek, MD, MME (Univ. Bern,) is a paediatrician, was one of the curriculum coordinators, Chairman of the Centre for Virtual Patients, and E-learning commissioner at Heidelberg Medical Faculty. He is now Head of the Department of Assessment and Evaluation at the Institute of Medical Education in Bern.

Bas de Leng, PhD, is an educational technologist at the Department of Educational Development and Research, Maastricht University, Maastricht, The Netherlands.

Andrzej A. Kononowicz, PhD, is Assistant Professor at Jagiellonian University Medical College in Krakow, Poland and post-doc researcher at Karolinska Institutet, Stockholm, Sweden.

Romy Kunzmann is a medical student at the Faculty of Medicine, Heidelberg University, Germany. She has been working on Virtual Patient evaluations at the Centre for Virtual Patients for more than two years.

Arno Muijtjens, MSc, PhD, statistician-methodologist, is Associate Professor at the Department of Educational Development and Research, Faculty of Health, Medicine, and Life Sciences, Maastricht University, The Netherlands.

Cees PM van der Vleuten, PhD, is Professor of Education, Chair of the Department of Educational Development and Research, Scientific Director of the School of Health Professions Education (SHE) at Maastricht University, Maastricht, the Netherlands.

Georg F. Hoffmann, M.D., Professor of Paediatrics and Chairman of the University Children's Hospital Heidelberg, Vice Dean of the Medical Faculty of the Ruprecht-Karls-University Heidelberg. He holds honorary appointments at the Medical Faculties of Padua, Italy, and Tongji, University Wuhan, China.

Burkhard Tönshoff, M.D., Ph.D., Professor of Paediatrics and Paediatric Nephrology, holds the position of a Vice Chairman of the Department of Paediatrics I, University Children's Hospital Heidelberg, Germany. He has been involved in the field of Virtual Patients for many years.

Diana H.J.M. Dolmans, PhD, is professor of Innovative Learning Arrangements within the School of Health Professions Education/Department of Educational Development and Research at Maastricht University, the Netherlands.

REFERENCES

- American Educational Research Association and American Psychological Association (1999): National Council on Measurement in Education. Standards for Educational and Psychological Testing. Washington, DC: American Educational Research Association.
- Balasubramaniam C, Poulton T, Huwendiek S (2009): Repurposing existing virtual patients; an Anglo-German case study. *Bio-Algorithms and Med-Systems* 5: 91-98
- Berman NB, Fall LH, Chessman AW, Dell MR, Lang VJ, Leong SL, Nixon LJ, Smith S (2011): A collaborative model for developing and maintaining virtual patients for medical education. *Med Teach* 33:319-24.
- Berman NB, Fall LH, Maloney CG, Levine DA (2008): Computer-assisted instruction in clinical education: a roadmap to increasing CAI implementation. *Adv Health Sci Educ Theory Pract* 13:373-83.
- Bowen JL (2006): Educational Strategies to Promote Clinical Diagnostic Reasoning. *N Engl J Med* 355: 2217-25
- Carifio J, Perla R (2008): Resolving the 50-year debate around using and misusing Likert scales. *Med Educ* 42:1150-1152.
- Cook DA, Beckman TJ (2006): Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med.* 119:166.e7-16. Review.
- Cook DA, Erwin PJ, Triola MM (2010): Computerized virtual patients in health professions education: a systematic review and meta-analysis. *Acad Med* 85:1589-602. Review.
- Cook DA, Triola MM (2009): Virtual patients: a critical literature review and proposed next steps. *Med Educ.* 43:303-11.
- Crick, J. E., & Brennan, R. L. (1983): Manual for GENOVA: A generalized analysis of variance system (ACT Technical Bulletin No. 43). Iowa City: American College Testing.
- Downing SM (2003): Validity: on meaningful interpretation of assessment data. *Med Educ.* 2003 Sep;37(9):830-7.
- Ellaway R, Poulton T, Fors U, McGee JB, Albright S (2008): Building a virtual patient commons. *Med teach.* 2008; 30, 170-174.
- Edelbring S, Broström O, Henriksson P, Vassiliou D, Spaak J, Dahlgren LO, Fors U, Zary N (2012): Integrating virtual patients into courses: follow-up seminars and perceived benefit. *Med Educ* 46: 417-425.
- Fall LH, Berman NB, Smith S, White CB, Woodhead JC, Olson AL (2005): Multi-institutional development and utilization of a computer-assisted learning program for the paediatrics clerkship: the CLIPP project. *Acad Med* 80:847-855.
- Gruppen LD & Frohna AZ (2002): Clinical Reasoning. In: Norman, Geoffrey R.; van der Vleuten C. P.M.; Newble, D.I. International handbook of research in medical education. Kluwer Academic Publishers. pp 205-230
- Huang G, Reynolds R, Candler C (2007): Virtual patient simulation at US and Canadian medical schools. *Acad Med.* 82: 446-451.
- Huwendiek S, Reichert F, Bosse HM, de Leng BA, van der Vleuten CP, Haag M, Hoffmann GF, Tönshoff B (2009b): Design principles for virtual patients: a focus group study among students. *Med Educ* 43: 580-8.
- Huwendiek S, De leng BA, Zary N, Fischer MR, Ruiz JG, Ellaway R (2009a): Towards a typology of virtual patients. *Med Teach.* 31:743-8.
- Huwendiek S & de Leng BA (2010): Virtual patient design and curricular integration evaluation toolkit. *Med Educ.* 44:519.
- Kim S, Phillips WR, Pinsky L, Brock D, Phillips K, Keary J (2006): A conceptual framework for developing teaching cases: a review and synthesis of literature across disciplines. *Med Educ* 40:867-876
- Mayring P. Qualitative Inhaltsanalyse - Grundlagen und Techniken. 11. Beltz Verlag, Weinheim und Basel; 2010.
- Poulton T, Conradi E, Fors U, Fischer M, Huwendiek S, Ellaway R, de Leng B, Davies D, Muntean V, Laidler P (2007): eViP: an EC-funded project to create a bank of multilingual, multicultural, virtual Patients. AMEE (Association for Medical Education in Europe) Conference 2007; 25.-29.08. 2007; Trondheim, Norway. Abstract 20/P14. p. 20.
- Poulton T, Balasubramaniam C (2011): Virtual patients: a year of change. *Med Teach.* 33:933-7.
- Shavelson R & Webb N (1991): Generalizability theory: A primer. Newbury Park, CA: Sage.

APPENDIX: PRIOR ORIGINAL ITEMS OF THE INSTRUMENT

The remaining items after the internal structure validation are presented in [blue](#)

Factor I: Authenticity of patient encounter and the consultation

1. While working on this case, I felt I had to make the same decisions a doctor would make in real life.
2. While working on this case, I felt as if I were the doctor caring for this patient.

Factor II: Cognitive strategies in the consultation

3. While working through this case, I was actively engaged in gathering the information (e.g. history questions, physical exams, lab tests) I needed to characterize the patient's problem.
4. While working through this case, I was actively engaged in revising my initial image of the patient's problem as new information became available.
5. While working through this case, I was actively engaged in creating a short summary of the patient's problem using medical terms.
6. While working through this case, I was actively engaged in thinking about which findings supported or refuted each diagnosis in my differential diagnosis.

Factor III: Coaching during consultation

7. I felt that the case was at the appropriate level of difficulty for my level of training.
8. The questions I was asked while working through this case were helpful in enhancing my diagnostic reasoning in this case.
9. The feedback I received was helpful in enhancing my diagnostic reasoning in this case.

Factor IV: Learning effect of consultation

10. After completing this case, I feel better prepared to confirm a diagnosis and exclude differential diagnoses in a real-life patient with this complaint.
11. After completing this case I feel better prepared to care for a real-life patient with this complaint.

Global score

12. Overall, working through this case was a worthwhile learning experience.

Open-ended questions

13. Special strengths of the case:
 14. Special weaknesses of the case:
 15. Any additional comments:
-

CHAPTER 5

Establishing guidance for Virtual Patients` implementation

Learner preferences regarding integrating, sequencing and aligning virtual patients with other activities in the undergraduate medical curriculum: A focus group study

Sören Huwendiek, Cecilia Dunker, Friedrich Reichert, Bas A. deLeng,
Diana Dolmans, Cees P. M. van der Vleuten, Martin Haag,
Georg Friedrich Hoffmann, Burkhard Tönshoff

Medical Teacher 2013; 35: 920–929

ABSTRACT

Context: E-learning resources, such as virtual patients (VPs), can be more effective when they are integrated in the curriculum. To gain insights that can inform guidelines for the curricular integration of VPs, we explored students' perceptions of scenarios with integrated and non-integrated VPs aimed at promoting clinical reasoning skills.

Methods: During their paediatric clerkship, 116 fifth-year medical students were given at least ten VPs embedded in eight integrated scenarios and as non-integrated add-ons. The scenarios differed in the sequencing and alignment of VPs and related educational activities, tutor involvement, number of VPs, relevance to assessment and involvement of real patients. We sought students' perceptions on the VP scenarios in focus group interviews with eight groups of 4–7 randomly selected students (n=39).

The interviews were recorded, transcribed and analysed qualitatively.

Results: The analysis resulted in six themes reflecting students' perceptions of important features for effective curricular integration of VPs: (i) continuous and stable online access, (ii) increasing complexity, adapted to students' knowledge, (iii) VP related workload offset by elimination of other activities, (iv) optimal sequencing (e.g.: lecture – 1 to 2 VP(s) – tutor-led small group discussion – real patient) and (v) optimal alignment of VPs and educational activities, (vi) inclusion of VP topics in assessment.

Conclusions: The themes appear to offer starting points for the development of a framework to guide the curricular integration of VPs. Their impact needs to be confirmed by studies using quantitative controlled designs.

INTRODUCTION

Contrary to earlier assumptions that e-learning resources, including virtual patients (VPs), would automatically reach their full potential when students were exposed to them, recent studies have reported low acceptance, usage and satisfaction among students when e-learning resources are poorly integrated with, or offered as an add-on to, other curricular components (Fischer et al., 2008; Haag et al., 2007). Research has also shown that VPs are particularly suited to foster clinical reasoning (Cook and Triola, 2009, Cook et al., 2010; Huwendiek et al., 2009b), and their use is on the increase in medical education (Leong et al., 2003; Fall et al., 2005; Huang et al., 2007; Berman et al. 2009; Cook and Triola, 2009, Cook et al., 2010; Huwendiek et al. 2009a). As part of our search for ways of optimizing the power of VPs to enhance medical education, the development of clinical reasoning skills in particular, we designed and studied integration scenarios for VPs, defined as “interactive computer simulation[s] of real-life clinical scenarios for the purpose of medical training, education, or assessment” (Ellaway et al., 2008). The present study is intended as a preliminary step towards the development of a guideline for the integration of VPs in medical curricula. In designing the study we drew on the (medical) education literature focusing on research of curricular integration of e-learning activities, including VPs, the sequencing of educational activities, simulation, blended learning and the development of clinical reasoning skills.

Although calls for studies of effective e-learning integration strategies in the 1990s (Friedmann, 1994) led to suggestions to set up a line of research to compare different integration scenarios (Cook et al., 2010; Friedmann, 1994; Cook, 2005, 2009; Triola et al. 2012), progress in this area has been limited (Cook et al., 2010; Cook, 2005, 2009; Triola et al., 2012). In 2007, the Association of American Medical Colleges (AAMC) called for instructional technology research focused on the integration of instructional technology into existing curricula (1) and educational settings (2) (AAMC, 2007). In our opinion the relevant question concerning VPs in this regard is which type of curricular integration is most likely to enhance clinical reasoning. Surveys on the integration of VPs into clinical clerkships have indicated improved usage, acceptance and perceived knowledge gain due to measures to stimulate curricular integration, such as VP-related student instructions, follow-up seminars, faculty development, formal assessment of VP content and avoidance of excess workload due to VPs by eliminating redundant teaching activities (Fischer et al., 2008; Berman et al., 2009; Edelbring et al., 2012).

The one study we found on sequencing in medical education concerned a comparison of different sequences of self-study (watching a PowerPoint presentation on the knee examination) and bedside teaching. Self-study followed by bedside teaching led to significantly better results on an OSCE and was preferred by 88% of the participating students (Hull et al., 2009). It is not clear, however, how these results would translate to other educational activities, such as VPs, lectures or small group discussions. In their book on sequencing in educational design, Ritter et al. proposed the following provi-

sional design rules related to sequencing (Ritter et al. 2007) in addition to the aspects of van Merriënboer and Sweller mentioned below (van Merriënboer & Sweller 2010). *Order* is especially important when there is a strong relationship among topics. The *amount of novelty and the amount to learn* should be maximized, adapted to learners' knowledge level. The amount of *time and repetitions* it takes to learn something and the spacing of practice should be considered. Types of problems should be varied to enhance *transfer of knowledge* to new problems. Compared to novices, expert learners should be given more freedom to (re-)order their learning activities.

A recent review of educational strategies to promote clinical reasoning (Bowen, 2006) emphasized techniques like summarizing the patient problem, reflecting, and critically appraising generated hypotheses, which seem to be applicable to VPs and other educational activities in an integration scenario (Huwendiek et al., 2009b).

Since VPs are simulations, albeit low fidelity ones, we also looked to simulation research for insights concerning the integration of VPs. A recent systematic review of simulation-based medical education research identified a need for ways to integrate simulation with other modalities which echoes the objectives pursued in the present study (McGaghie et al., 2010). Similarly, a best evidence medical education review (Issenberg et al., 2005) advocated that simulation be integrated in the overall curriculum and learners should practise at increasing levels of difficulty. This recommendation might be translated to the scheduling of VPs (low complexity and fidelity) before real patient encounters (high complexity and fidelity). This approach appears to resonate with curriculum design principles proposed by Van Merriënboer and Sweller (van Merriënboer and Sweller, 2010), who recommend a similar sequencing of learning tasks from simple to complex and from low-fidelity to high-fidelity environments.

Research of blended learning addresses issues that are akin to integration and sequencing, i.e. striving for the most advantageous combinations of face-to-face instruction and computer-mediated instruction to maximize the strengths and minimize the weaknesses of each instructional method (Bunk & Graham 2006). This might imply a preference for the use of VPs for self-study and face-to-face sessions for real patient encounters with direct feedback from supervisors.

So, while the literature does offer some general suggestions concerning a general guiding framework for curricular integration of VPs, it offers no concrete guidance. We therefore designed a study to explore aspects of the curricular integration of VPs. Based on the literature and our own experiences with VPs (Fischer et al., 2008; Huwendiek et al., 2009a, b, c), we exposed students to different educational scenarios with VPs, which differed in the sequencing and alignment of VPs and other educational formats, tutor involvement, number of VPs, relevance of VPs to assessment and the involvement of real patients. As students have first-hand experience of educational activities, we explored their perceptions of the scenarios and of the optimal VP scenario for fostering clinical reasoning skills.

METHODS

Setting

Every week of the four-week paediatric clerkship in our institution, students attend daily 1.5 hour symptom-oriented lectures, three two-hour bedside teaching sessions, one 90-minute problem-based learning tutorial with paper cases and a tutor-led small group discussion on VPs. Elsewhere we have described how the VPs were designed (Huwendiek et al., 2009b), and appendix 1 describes one of the VPs used in the present study to illustrate its design based on the VP design typology we proposed in an earlier study (Huwendiek et al., 2009a).

During the paediatric rotation, students were exposed to nine educational scenarios with VPs. The VPs were integrated with other activities in eight scenarios and an unrelated add-on in one scenario (Figure 1). All scenarios were aimed at fostering clinical reasoning skills. The scenarios differed in topic and the number and sequencing of activities. For example, one scenario addressed the leading symptom of vomiting, and the educational activities consisted of an interactive lecture, instructions for VP-guided self-study - which students could undertake on their own or with others -, learning objectives for self-study, a tutor-led small group discussion and bedside teaching. Students received detailed information about the integration of VPs from teachers, hand-outs and electronic schedules, which included learning objectives of educational activities and web links to VPs. The teachers who participated in the study (n=11) attended workshops on (a) facilitating small group discussions with a focus on fostering clinical reasoning (Bowen 2006) and (b) working with VPs, (c) integrating VPs into the clerkship and (d) informing students about the integration of VPs and learning objectives. Students could discuss VP-related issues on an online forum, and the integrated VPs were relevant to the paediatric course exam, which consisted of VP key feature cases on eight relevant paediatric leading symptoms (Huwendiek et al. 2009c). The students received no credits for the work-up of the integrated VPs.

The non-integrated VPs (n=15) were offered as self-study tasks but without explicit linkage to other educational activities. There was an online forum to discuss the VPs, but the cases had no relevance to the exam. Students received credits for these VP assignments.

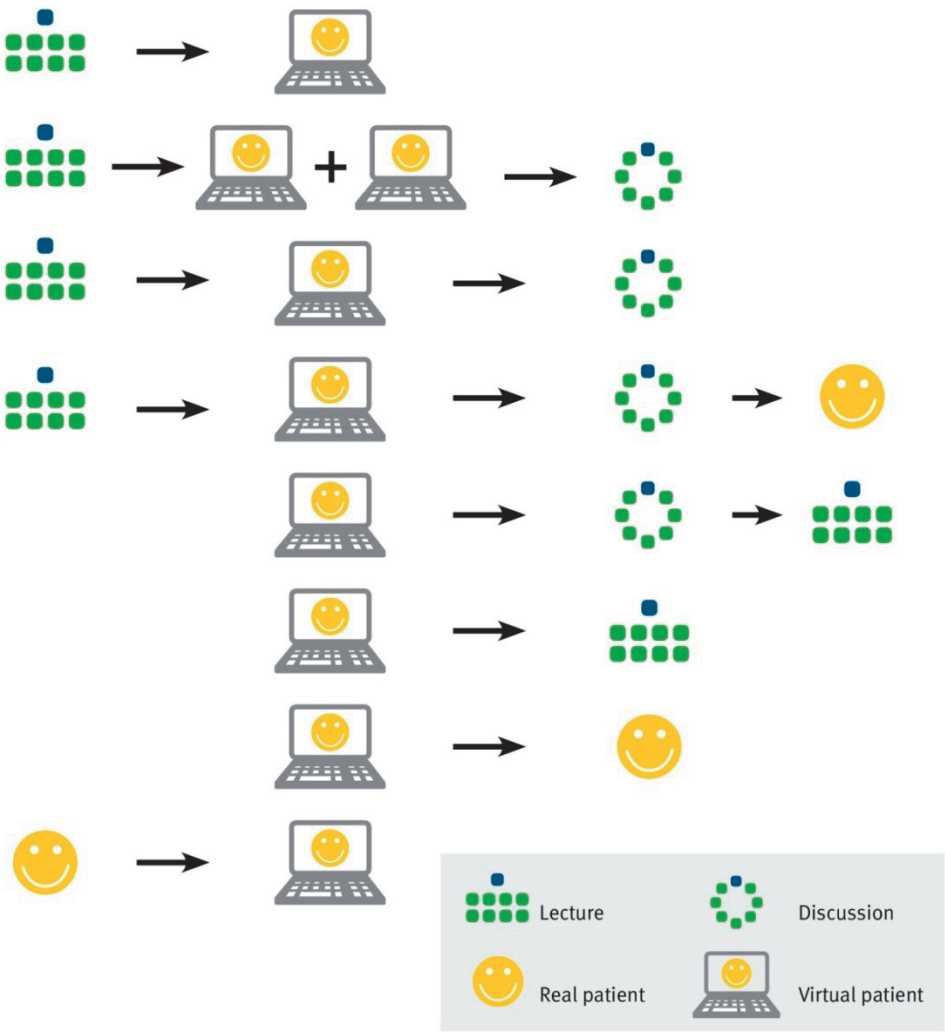


Figure 1: VP Integration scenarios 1-8

Subjects

Three groups of approximately forty students (N=116) that were exposed to the nine VP scenarios in varying order received an invitation to participate in a study about clinical education. To avoid bias, additional information about the study was provided to the participating students only. Of the 65 (56%) students who volunteered to participate, 39 were randomly selected to attend one of eight focus group interviews (4-7 students). The participants received a small financial incentive.

Focus group discussion and analysis

We explored students' perceptions of the VP scenarios in focus group interviews, because this qualitative approach is appropriate for exploring participants' views as well as the notions and considerations underlying these views (Krueger et al., 2000). Interviews of peer groups provide a relatively safe environment by reducing the imbalance of power between researcher and subjects and promoting in-depth discussion, which may lead to refinement or modification of views. To counteract the potential drawback of social pressure from peers or moderators inhibiting the open expression of opinions, the moderator encouraged all participants to contribute, emphasized the value of differing opinions and views and asked the participants to write down their thoughts on a topic before group discussion started. Eight two-hour focus group sessions were held on different days, moderated by one of the authors (SH), who is an experienced moderator of small groups. Consistency across group interviews was promoted by a questioning route (Krueger and Casey, 2000) (Table 1).

Table 1: The questioning route used by the moderators in guiding the focus group discussions

The VPs you were given in the previous module were linked or not linked to other teaching events in the module.
1.Which aspects of the integration of VPs did you find the most helpful for improving your clinical reasoning skills?
2.Which aspects of the integration of VPs did you find the least helpful for improving your clinical reasoning skills?
3.Let us look at scenario 1 (short description of scenario): a) Which factors promoted your clinical reasoning skills? b) Which factors were not helpful in promoting clinical reasoning skills?
4.The same questions for all nine scenarios.
5.What should a single scenario look like to best promote clinical reasoning and to best prepare you (students) for real patients?
6.What does an ideal overall plan for the integration of VPs in the paediatrics module look like to foster clinical reasoning?
7.How did the relevance of the cases to the exam influence your perception of/work on the VPs? Should VP content be included in the exams? If yes, why?

Assistant moderators (CD, FR) took comprehensive notes and recorded the sessions on video. The recordings were transcribed literally, and summary reports were sent to the participating students for approval.

Three of the authors (CD, FR, SH) analysed the transcripts. In accordance with guidelines for qualitative content analysis, they first read all the transcripts while identifying and highlighting preliminary themes (Mayring 2010). Next, they established themes and subthemes in an iterative process in which coded themes were discussed by the research team and the discussions, in turn, informed the coding process. The process continued until consensus was reached. All students received a summary report of their

session with illustrative quotations. Whenever possible, the strength of an opinion and the frequency of expressed group support were indicated. All participants approved the summary reports and there was a high level of consistency across the groups. As no significant new ideas emerged during the eighth session, no further sessions were organized (Krueger and Casey, 2000).

In the country where the study was carried out, ethical approval is not required for this type of educational study. However, we confirm that participation was voluntary, participants cannot be identified from the material presented and no plausible harm to participating individuals can arise from the study.

RESULTS

The analysis resulted in six main themes - some with subthemes - which reflect students' perceptions and preferences with regard to optimal curricular integration of VPs. Representative citations from the interviews are provided for each theme with numbers in parentheses indicating the focus group and the student. Additional citations can be found in Table 2.

1. Continuous and stable online access to VPs

Students valued continuity and stability of access to VPs, e.g. via a learning platform, because it created flexibility, allowing them to study at convenient times and places and revisit VPs when preparing for the exam.

"... stable online access to the VPs means they can be worked through whenever this is convenient, e.g. at home. This is especially helpful in preparing for the exam." (1,2)

2. Increasing complexity of VPs adapted to students' knowledge level

According to students, early in the clerkship when students have limited knowledge, VPs should be relatively easy, and increase in complexity as students' knowledge and experience increase during the rotation.

"... the level of difficulty of the VPs should depend on students' current level of knowledge and how many weeks of paediatric lectures and bedside teaching they have attended." (1,1)

3. Excess workload due to the introduction of VPs should be offset by elimination of existing curricular activities.

It was important for students that the additional workload due to the integration of VP tasks in the clerkship was offset by removal of existing clerkships activities - but not those involving real patients. Students appreciated dedicated time for self-study with VPs. It was suggested to make some VP activities obligatory, especially at first to stimulate students to familiarize themselves with this learning resource.

„The cases should not be an add-on to the course, but an integral part of it and a substitute for other teaching sessions.“ (3,2)

4. Optimal sequencing of VP work and related education activities

4a Lecture before VP work

Students preferred VP work to be scheduled after the lecture, as it enabled them to apply knowledge from the lecture to the VP. This boosted students' confidence and helped them remember important aspects.

„Working through the VP before attending the lecture was not as useful to me as attending the lecture before doing the VP. I feel I was able to get more out of the VP when I first attended the lecture in which the substance and procedures were explained.“ (5,2)

4b VP work before tutor-led small group discussion focused on clinical reasoning

Students preferred a tutor-led small group session to be scheduled after VP self-study, because they took self-study more seriously when they knew they were likely to be questioned about the VP. According to students the discussion improved their understanding of clinical reasoning, especially for difficult and complicated cases and early in the clerkship. Wrap-up discussions in which two VPs with relevant differential diagnoses were compared improved students' understanding of disease and helped them recognize key diagnostic features.

“... I learn well with patients. That was also the case for the VPs. Working through VPs helps me to remember cases ... And afterwards during the group discussion, the differences between the cases were highlighted and discussed. The cases were compared and contrasted and the differences highlighted. That was very helpful.“ (6,4)

4c VP work before real patient encounter

According to students, it was highly meaningful, motivating and supportive of their learning to do a VP before an encounter with a real patient (bedside teaching) with similar symptoms. Having studied the VP made students feel well prepared and confi-

dent to engage in these encounters, because they knew which questions to ask, which exams to perform and which differential diagnoses and confirming and disconfirming features to consider. This enabled students to focus on the execution of tasks (how to ask history questions, perform a physical exam) instead of on the selection of tasks (which questions, which exams etc.).

“I saw a child on ward 1 with meningitis and the child had the same symptoms about which questions were asked in the related VP case: which symptoms would you look for during the physical examination, and which questions are especially important in the medical history... Since I had done the VP, I did not have much difficulty figuring out the diagnosis. I really thought that having done the VP first had optimally prepared me for the real patient...” (1,1)

4d Optimal sequence: 1. lecture – 2. 1-2 VP(s) – 3. tutor-led small group discussion – 4. real patient

Students felt that the sequence lecture – 1-2 VP(s) - small group - real patient offered maximum support for their learning and optimal preparation for the real patient encounter at the end (see also figure 2). Students thought that lectures should not be compulsory, as they preferred to decide for themselves whether they obtained relevant basic knowledge from lectures or from resources like books, clinical guidelines, e-learning modules etc.

„In my opinion the sequence - lecture, 1-2 VPs, small group discussion, real patient – is the best alternative. During a VP task we focus on a specific topic, consider the differential diagnosis and review our knowledge. We can use the knowledge from the lecture too... So in all I think it’s better to practise with the VPs first, find out how much you already know and where more work is needed. Then in the small group discussion it is good to review the main issues and finally, at bedside teaching, we know what needs to be done. This approach makes us feel optimally prepared.” (5,2)

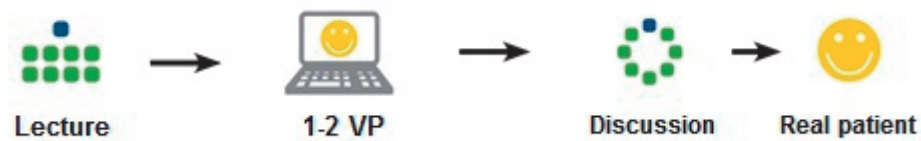


Figure 2: Optimal sequence (4 d)

5. Optimal alignment of VP work and related education activities

5a Alignment of content

Students thought that VPs and related educational activities should be complementary in terms of content. Activities should build on each other and allow students to apply newly obtained knowledge. For example, in a VP task, students can apply knowledge from the preceding lecture. Students indicated that the group discussion need not be limited to differential diagnosis but could also cover relevant clinical cases from the tutor's experience.

"I worked through a VP, and then I went to the oncology ward where I saw a patient with a similar disease. After that we discussed the disease. It was great that it was all so well coordinated and it added depth and some much needed repetition to the case." (7,4). "..."So the order should be: lecture, VP, small group discussion, real patient and the contents of the sessions should match" (6,4)

5b Alignment of instructional methods

According to students, educational modalities need methods that optimize learning. VPs, for example, were considered to be excellent for self-study, because they were more suited to offer challenges and formative evaluation and feedback than any of the other modalities. Face-to-face discussion sessions should not revisit the preceding VPs but use methods to foster clinical reasoning in an interactive small group setting. Students preferred lectures to be symptom oriented and interactive, while in small group sessions they looked for a tutor who actively challenged them when discussing the differential diagnosis including disconfirming and confirming factors and feedback. Student preferences with regard to bedside teaching involved being allowed to do the history and physical examination, supervised by teachers who challenge them with questions about clinical reasoning. According to students, it is important to generate concrete and clear learning objectives to be pursued during self-study between different activities in order to ensure continuity of the evolving learning process.

"I found the following most useful for learning: That the VP of the self-study assignment came after the interactive lecture about the cardinal symptom in question and that we could discuss the most important concepts and ask questions during the small group discussion session. We then saw a real patient and took a history and performed the required clinical examinations. The tutor who was in charge of the bedside teaching gave us feedback on our performance. In this way, I was able to make the most of these sessions and learn a lot from them." (2,2)

5c Information for students

Students attached great importance to information from teachers about how VPs were integrated in the curriculum and aligned with other educational activities. They recom-

mended that a list of VPs and related educational activities be provided in a hand-out and/or online timetable, with links to related VPs in the information about an educational activity. Knowing how VPs were linked to other activities motivated students to take self-study more seriously and consequently improved their learning.

“...The teachers should indicate during sessions which sessions are connected. So that one always knows in advance that the educational sessions deal with the same topic, from the lecture to the real patient.” (2,4)

5d Trained teachers

Students considered it important for teachers to be well informed about the substance and curricular integration of the VPs. They also expected teachers to be experienced clinicians and have skills to facilitate small group discussions on clinical reasoning, for example contrasting different cases, quizzing students about confirming and disconfirming features of differential diagnoses and highlighting relevant clinical aspects. The learning gains from tutor-led wrap-up sessions depended strongly on the tutor's facilitating skills and clinical knowledge.

“How valuable the small group discussions are depends very much on the ability of the tutors to facilitate relevant discussions about differential diagnoses. For example, whether the tutor questions students about what would support a certain diagnosis. It also depends on how difficult the VP cases are and on the clinical experience of the tutor.” (1,5)

6. The relevance of VPs to assessment

Students supported their plea for including VPs in the course exam by the following argumentation. It is important for students to study VP tasks because this enhances learning and clinical reasoning skills, which are important for students' future medical practice. Students are motivated to study VPs that are relevant to their future profession AND the course exam. Students felt that if the content of VP cases was assessed, they were stimulated to study more thoroughly, more attentively and more frequently.

“The cases should be relevant to the course exam, because these cases prepare us optimally for our future roles and their relevance to the exam motivates us to spend more time on relevant topics.” (4,3)

Table 2: Additional illustrative citations from focus group discussions relating to VP integration. The numbers in parentheses indicate focus group and student, respectively.

1. Continuous and stable online access to VPs

"... if you can work through a VP at home, you can check your knowledge about a certain topic by working through the relevant VP to see how you would do in a more realistic situation." (1,6)

3. VP work balanced with elimination of other curricular activities

"It is great that VPs are made available to us as part of the paediatric programme. It offers us a better selection of case topics from which we can choose topics that interest us, but personally I feel that they should not replace real patients." (2,4)

"I think it makes sense to make at least the first VP compulsory, as this would allow every student to get acquainted with the programme and work with it. Afterwards they can decide for themselves whether they consider the VPs helpful or not." (1/1)

4. Optimal sequencing of VP work and related educational activities

4a Lecture before VP work

"In my opinion, the best way to do it is to have the lecture first to get an overview of the topic and then work through the cases independently. This means that you know which differential diagnoses to consider and which diagnostic tools might be useful." (6,2)

"It is very helpful when the lecture does what books cannot do, namely discuss the characteristic symptoms of diseases and the general approach for managing diseases. In this way I have a list of differential diagnoses ready in my head, which I can use for example to answer the question at the beginning of the VP. (6,5)

4b VP before tutor-led small group discussion focused on clinical reasoning

"Besides learning through the VP it was primarily through the small group discussion sessions that I learned to think in terms of differential diagnoses. Also, it is not sufficient for us to learn about a disease or a case; we also need to learn about the best management of patients with a specific cardinal symptom. The small group discussions really helped me to do just that and that's why I found them very helpful." (4,4)

"I found the small group discussion sessions really useful. Students prepare more thoroughly if they know that the cases will be discussed with a tutor. Important questions that come up and would normally not be addressed without a tutor can be dealt with. Furthermore, it is easier to remember important content than would be the case otherwise." (5,2)

"How valuable the small group discussion is depends also on the level of difficulty of the VP." (1,5)

4c VP before real patient encounter

"My experiences were as follows: After doing a VPs, I met and interviewed a real child on the ward in three instances. So for example in the morning I did a VP with poststreptococcal glomerulonephritis, and then in the afternoon I went to the ward where I saw a child with poststreptococcal glomerulonephritis. So it worked out really well. I already had a couple of differential diagnoses in my head and knew what not to include. Having worked through the corresponding VP beforehand meant that I had a much better understanding of the underlying pathology and that prevented me from asking many less useful questions. This also meant I could concentrate on how to ask the question and not just on what I questions I needed to ask. That was a moment of success for me because I felt like I did well. Furthermore, seeing a child during bedside teaching motivated me to study the child's disease in more detail; I actually sat down at home and read up on it. That was extremely motivating as well." (3,5)

4d Optimal sequence: lecture – 1-2 VP(s) – tutor-led small group discussion – real patient

"In my opinion, a VP is more accurate and complex than the lecture. Therefore I think it makes sense to start with the easier task and move on to more challenging tasks rather than the other way around." (6,4)

5. Optimal alignment of VP work and corresponding educational activities

5a Alignment of content

"When I have worked through a VP and then see a similar patient on the ward, ... I can ask questions about the case and discuss topics which were left unanswered after the VP." (1,6)

5b Alignment of instructional methodology

"What I found especially helpful about the small group discussion was when the tutor quizzed us about the cases until we could explain these ourselves. If someone tells me the correct answer, I might remember what he said after 5 seconds, but if I really have to think long and hard about a question, then I would most likely still be able to come up with the in 6 months' time." (3,5)

"I found the small group discussion sessions really useful because it made us use more than one of our senses to aid learning. During the discussion we had the opportunity not only to read or listen but also to speak and explain important concepts. This enriched my learning experience considerably and opened a new channel to promote understanding." (3,3)

"What I found helpful about the online timetable is that it clearly stated what we had to study and what we were required to cover for each session. Furthermore, I found it especially useful that the relevant pages and the learning objectives for a teaching session were clearly listed. (3,5)

5c Information for students

"It motivated me in my studies, when we were told in advance that a certain lecture was related to a certain VP and small group session." (2,5)

„When the lecturer clearly pointed out during the lecture: pay attention, you will need this information for the VP case, that was really helpful." (2,4)

5d Trained teachers

"It would be nice if we were able to instantly recognise which educational sessions are connected, including lectures and sessions with real patients. The tutors should be aware of this and point it out to the students." (8,4)

"It really helps me when we go over the case again in a structured manner during the small group session. The tutors expected us to present the VPs to the group. Making concluding remarks about the case, highlighting what was especially important about it and getting helpful feedback on our performance has definitely helped me to enjoy a more fruitful learning experience." (2,4)

6. Relevance of VPs to assessment

"Naturally, when I knew the contents of VP cases were likely to come up in the exam, I went over the cases a second time. I wrote out explicitly what the cardinal symptoms were, what the typical differential diagnoses were, what diagnostic investigations I ought to perform, what symptoms I needed to watch out for and how to plan the treatment. This is how I made my notes for every VP." (4,2)

"If I had known they were not at all or less relevant to the exam, then I wouldn't have worked through the VPs in such great depth." (3,4)

DISCUSSION AND CONCLUSION

In our search for evidence to support guidelines for the curricular integration of VPs, we explored students' perceptions of different scenarios aimed at developing clinical reasoning skills in which VPs were integrated with or only added to educational activities. The analysis of eight focus group interviews suggested six themes which students considered important for the optimal integration of VPs.

The themes reveal students' preferences regarding the sequencing of e-learning and other educational activities (theme 4). Interestingly, these preferences are consistent with results of research on sequencing showing that e-learning should precede real patient encounters (Hull et al., 2009). Theme 4 also reflects the students' preference for lectures before VPs (theme 4a), and VP work before small group discussions (theme 4b), while lecture–1-2 VPs–tutor-led small group discussion–real patient encounter is the sequence of preference (theme 4d, figure 2). The reasons students gave for their sequence preferences appear to be in line with emerging curriculum design principles, such as “from simple to complex” (themes 2 and 4d) and “from low to high fidelity” (Van Merriënboer and Sweller, 2010) (theme 4d). Similarly, meta-analyses of simulation research have demonstrated the importance of steadily increasing complexity adapted to learners' abilities (Issenberg et al., 2005; McGaghie et al., 2010) (themes 2 and 4d), and conclusions on sequencing from general education research also resonate with students' views on how to order learning tasks (Ritter et al. 2007).

The analysis revealed themes based on student perceptions which could provide preliminary guidance not only on the sequencing but also on the alignment of VPs and other educational formats commonly used to foster clinical reasoning. It should be noted, however, that the themes are supported by students' views only, and need substantiation by further research. It appears to be important that VPs and related educational activities should use appropriate instructional methods and address aspects to which they are particularly suited. Furthermore teachers should be properly trained for their tasks (theme 5). Strategies to foster clinical reasoning (Bowen, 2006) should be implemented both in VPs (Huwendiek et al., 2009b) and in related educational activities. Sequencing and alignment of VPs and other educational events (themes 4 and 5) were considered to be especially helpful in fostering clinical reasoning and preparing students optimally for clinical reasoning in the treatment of real patients.

Theme 6 recommends that VP-related learning objectives should be relevant to assessment. While this is in line with general principles of learning and assessment (Larsen et al., 2008; Schuwirth et al., 2011) and a previous article on VP integration (Fischer et al., 2008), it also underscores the practical relevance of VPs. The findings of our study are in line with studies indicating that faculty development, formal assessment of VP content and balancing of VP use with elimination of redundant teaching activities is important (Fischer et al., 2008; Berman et al., 2009; Edelbring et al., 2012). It should be noted that the students cautioned that VPs must never replace learning with real patients.

To our knowledge, there are no other translational studies (Ringsted et al., 2011) that have endeavoured to shed light on optimal strategies for integrating VPs into clinical clerkships by exposing medical students to different curricular integration scenarios with VPs, informed by the education literature. The present study may be the first one to respond to the long-standing call for comparative studies of different integration scenarios for e-learning, although it should be noted that the comparison is exploratory

relying solely on interview data (Cook et al., 2010; Friedmann, 1994; Cook, 2005, 2009; Triola et al., 2012; AAMC, 2007). Another new feature of this study is that it suggests ways to sequence and align VPs and other teaching modalities. The results give indications that VPs designed in accordance with specific principles (Huwendiek et al., 2009b) and integrated in the curriculum in accordance with the notions reflected in the six themes are perceived by students as offering not only excellent support for their learning but also excellent preparation for clinical reasoning in real patient encounters. Medical teachers and curriculum designers could use the recommendations that can be derived from the themes as a preliminary guideline on the integration of VPs into the curriculum. It is also important, however, that the themes are supported by evidence from further studies.

The qualitative paradigm we employed did not require the measurement of quantitative outcomes. However, our findings, and other models related to VP integration and sequencing, would merit confirmation in controlled experiments using outcomes of knowledge, skills and behaviours. A limitation is that focus group interviews are susceptible to bias, such as that due to the moderator's influence, although we took measures to prevent this. Since the abundance of options for integrating VPs in the curriculum precluded inclusion of all possible scenarios in the study, some relevant aspects may have been overlooked. Furthermore, as the study was conducted among undergraduate medical students in one institution and with one type of VP, the generalizability of the results is inevitably limited.

ACKNOWLEDGEMENTS:

The authors would like to thank all students who participated in this study and Mereke Gorsira for critically reading and correcting the English manuscript.

DECLARATION OF INTERESTS:

The authors report no conflict of interest. The authors alone are responsible for the content and writing of this article.

PRACTICE POINTS:

- The analysis of focus group discussions yielded themes concerning curricular integration of VPs to foster clinical reasoning. While requiring confirmation by further studies, these themes can offer preliminary guidance to educators:

- Additional workload arising from VP tasks is to be offset by the elimination of other curricular activities.
- The preferred sequence of VPs and related educational activities is: lecture – 1-2 VP(s) – tutor-led small group discussion – real patient.
- VP work and related education activities should be optimally aligned.
- VP topics should be formally assessed.

GLOSSARY :

Blended Learning: Combining face to face with technology mediated learning environments. The challenge is to find blends that take advantage of the strengths of each environment and avoid the weaknesses. Bonk CJ, Graham CR: Handbook of blended learning: global perspectives, local designs. San Francisco: Pfeiffer Publishing; 2006.

Virtual Patients: Interactive computer simulation[s] of real-life clinical scenarios for the purpose of medical training, education, or assessment. Reference: Ellaway R, Poulton T, Fors U, McGee JB, Albright S. Building a virtual patient commons. Med Teach. 2008; 30:170–174.

NOTES ON CONTRIBUTORS:

Sören Huwendiek, MD, MME (Univ. Bern,) is a Paediatrician, was one of the Curriculum Co-ordinators, Chairman of the Centre for Virtual Patients and E-learning Commissioner at Heidelberg Medical Faculty. He is now Head of the Department of Assessment and Evaluation at the Institute of Medical Education in Bern.

Friedrich Reichert, MD, Paediatric Resident at Olgahospital, Klinikum Stuttgart, wrote his MD-Thesis about the design of Virtual Patients and their integration into medical curricula.

Cecilia Duncker, MD, Paediatric Resident at the University Hospital in Kiel wrote her MD-Thesis about the optimal integration of Virtual Patients into medical curricula.

Bas de Leng, PhD, is an educational technologist at the Department of Educational Development and Research, Maastricht University, Maastricht, The Netherlands.

Diana H.J.M. Dolmans, PhD, is professor of Innovative Learning Arrangements within the School of Health Professions Education/Department of Educational Development and Research at Maastricht University, the Netherlands.

Cees PM van der Vleuten, PhD, is Professor of Education, Chair of the Department of Educational Development and Research, Scientific Director of the School of Health Professions Education (SHE) at Maastricht University, Maastricht, The Netherlands. He holds honorary appointments in the University of Copenhagen (Denmark), King Saud University (Riyadh) and Radboud University (Nijmegen).

Martin Haag, PHD is a Professor of Software Engineering and Dean at the Faculty of Informatics at Heilbronn University, Germany. Furthermore he's head of the Centre for Virtual Patients at the University of Heidelberg, Germany and speaker of the working group technology enhanced learning of the German Society for Medical Informatics.

Georg F. Hoffmann, M.D., Professor of Pediatrics and Chairman of the University Children's Hospital Heidelberg, Vice Dean of the Medical Faculty of the Ruprecht-Karls-University Heidelberg. He holds honorary appointments at the Medical Faculties of Padua, Italy, and Tongji, University Wuhan, China.

Burkhard Tönshoff, M.D., Ph.D., Professor of Pediatrics and Pediatric Nephrology, holds the position of a Vice Chairman of the Department of Pediatrics I, University Children's Hospital Heidelberg, Germany. He is involved since many years in the field of Virtual Patients.

REFERENCES

- AAMC. Effective Use of Educational Technology in Medical Education. Colloquium on Educational Technology: Recommendations and Guidelines for Medical Educators. Washington, DC: AAMC Institute for Improving Medical Education; 2007.
- Bonk CJ, Graham CR: Handbook of blended learning: global perspectives, local designs. San Francisco: Pfeiffer Publishing; 2006
- Berman N, Fall LH, Smith S, Levine DA, Maloney CG, Potts M, Siegel B, Foster-Johnson L. Integration strategies for using virtual patients in clinical clerkships. *Acad Med.* 2009;84:942-9.
- Bowen JL. Medical education: Educational strategies to promote clinical diagnostic reasoning. *N Engl J Med.* 2006;355:2217-2225.
- Cook DA. The research we still are not doing: an agenda for the study of computer-based learning. *Acad Med.* 2005;80:541-548.
- Cook DA, Triola MM: Virtual patients: a critical literature review and proposed next steps. *Med Educ.* 2009;43:303-11.
- Cook DA. The failure of e-learning research to inform educational practice, and what we can do about it. *Med Teach.* 2009;31:158-62.
- Cook DA, Erwin PJ, Triola MM: Computerized virtual patients in health professions education: a systematic review and meta-analysis. *Acad Med.* 2010;85:1589-602.
- Edelbring S, Broström O, Henriksson P, Vassiliou D, Spaak J, Dahlgren LO, Fors U, Zary N: Integrating virtual patients into courses: follow-up seminars and perceived benefit. *Med Educ* 2012; 46: 417–425.
- Ellaway R, Poulton T, Fors U, McGee JB, Albright S. Building a virtual patient commons. *Med teach.* 2008; 30:170-174.
- Fall LH, Berman NB, Smith S, White CB, Woodhead JC, Olson AL. Multi-institutional development and utilization of a computer-assisted learning program for the pediatrics clerkship: the CLIPP Project. *Acad Med.* 2005; 80:847-855.
- Fischer MR, Hege I, Hörnlein A, Puppe F, Tönshoff B, Huwendiek S: [Virtual patients in medical education: a comparison of various strategies for curricular integration]. *Z Evid Fortbild Qual Gesundheitswes.* 2008;102: 648-53.
- Friedmann CP. The research we should be doing. *Acad Med* 1994, 69: 455-457
- Haag M, Singer R, Bauch M, Heid J, Hess F, Leven FJ. Challenges and perspectives of computer-assisted instruction in medical education: Lessons learned from seven years of experience with the CAMPUS system. *Methods Inf Med.* 2007;46: 67– 69.
- Huang G, Reynolds R, Candler C. Virtual patient simulation at US and Canadian medical schools. *Acad Med.* 2007;82: 446-451.
- Hull P, Chaudry A, Prasthofer A, Pattison G. Optimal sequencing of bedside teaching and computer-based learning: a randomised trial. *Med Educ.* 2009;43:108-12.
- Huwendiek S, De leng BA, Zary N, Fischer MR, Ruiz JG, Ellaway R. Towards a typology of virtual patients. *Med Teach.* 2009a ;31:743-8.
- Huwendiek S, Reichert F, Bosse HM, de Leng BA, van der Vleuten CP, Haag M, Hoffmann GF, Tönshoff B. Design principles for virtual patients: a focus group study among students. *Med Educ.* 2009b ;43: 580-8.
- Huwendiek S, Hanebeck B, Bosse HM, Haag M, Hoffmann GF, Tönshoff B. Lernen und Prüfen mit virtuellen Patienten am Zentrum für Kinder- und Jugendmedizin des Universitätsklinikums Heidelberg: Ergebnisse der Evaluation im Rahmen des E-Learning-Preises Baden-Württemberg 2007. *GMS Med Inform Biom Epidemiol.* 2009c;5:Doc10. Retrieved October 1 from urn:nbn:de:0183-mibe0000896
- Issenberg SB, McGaghie WC, Petrusa ER, Lee Gordon D, Scalese RJ. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Med Teach.* 2005, 27:10-28.
- Krueger RA, Casey MA. Focus groups: a practical guide for applied research. 3 ed. London Sage Publications, Inc; 2000.

CHAPTER 5

- Larsen DP, Butler AC, Roediger HL 3rd.. Test-enhanced learning in medical education. *Med Educ.* 2008;42:959-66.
- Leong SL, Baldwin CD, Adelman AM. Integrating Web-based computer cases into a required clerkship: development and evaluation. *Acad Med.* 2003;78:295-301.
- Mayring P. *Qualitative Inhaltsanalyse - Grundlagen und Techniken.* 11. Beltz Verlag, Weinheim und Basel; 2010.
- McGaghie WC, Issenberg SB, Petrusa ER, Scalese RJ. A critical review of simulation-based medical education research: 2003-2009. *Med Educ.* 2010; 44:50-63.
- van Merriënboer JJ, Sweller J. Cognitive load theory in health professional education: design principles and strategies. *Med Educ.* 2010; 44:85-93.
- Ringsted C, Hodges B, Scherpbier A. 'The research compass': an introduction to research in medical education: AMEE Guide no. 56. *Med Teach.* 2011;33:695-709.
- Ritter, F E, Nerb, J, Lehtinen, E and O'Shea, T M, Eds. (2007). *In Order to Learn: how the sequence of topics influences learning.* New York, Oxford University Press.
- Schuwirth LW, Van der Vleuten CP. Programmatic assessment: From assessment of learning to assessment for learning. *Med Teach.* 2011;33: 478-85.
- Triola M, Huwendiek S, Levinson AJ, Cook DA: *New Directions in e-Learning Research in Health Professions Education: Report of Two Symposia.* *Med Teach.* 2012;34:e15-20.

APPENDIX 1:

Description of one representative VP using the proposed VP design typology (Huwendiek et al. 2009a)

Title	Infant with fever
Description	An infant with fever presents to the paediatric emergency room. The learner is the doctor in charge of this patient and has to take the same decisions a doctor would need to make in real life (e.g. choosing history questions, physical exams, lab tests, DD, therapies, etc.).
Language	German
Identifier	evip:vp: 1000263
Provenance	Huwendiek, Höcker, Seidel, University Childrens`s Hospital Heidelberg
Typical study time	45 minutes
Educational level	Undergraduate – final years
Educational modes	Learning and formative assessment
Coverage	Paediatrics
Objectives and outcomes	Learners will learn how to gather relevant information, diagnose, treat and manage an infant presenting with fever. At the end of the activity learners will be able to: (1) Describe how to make a diagnosis in an infant with fever without focus; (2) Describe how to rule out differential diagnoses; (3) Describe how to manage an infant with fever.
Path type	Linear string of pearls
User modality	The single user plays the role of the paediatrician in charge, first in the emergency room then on the ward
Media and resources	Text, static graphics, videos
Narrative use and patient focus	Told from the perspective of the patient and the patient's mother, active.
Interactivity use	Long menu questions Multiple choice questions Free text questions Overall number of cognitive interactions: 133
Feedback use	Feedback is given on each decision by comparison with the decision made by an expert. In many cases reasons are given to explain why a decision is right or wrong. Feedback is provided immediately and statistically summarized at the end of the activity.
Originating system	CAMPUS classic
Format	Interactive Java-Applet inside Web-Browser
Integration and dependence	Java Plug-in respectively Java Runtime Environment (JRE), CAMPUS backend, otherwise independent

CHAPTER 6

Assessing Virtual Patients` implementation

Electronic assessment of clinical reasoning in clerkships: A mixed-method comparison of long menu key feature problems with context rich single-best-answer questions

Sören Huwendiek, Friedrich Reichert, Cecilia Duncker, Bas A. deLeng, Cees P.M. van der Vleuten, Arno Muijtjens, Hans-Martin Bosse, Martin Haag, Georg Friedrich Hoffmann, Burkhard Tönshoff, Diana Dolmans

Manuscript under review

ABSTRACT

Background: It remains unclear which item format would best suit the assessment of clinical reasoning: context-rich single best answer questions (crSBAs) or key-feature problems (KFPs). This study compared KFPs and crSBAs with respect to students' acceptance, their educational impact and psychometric characteristics when used in a summative end-of-clinical-clerkship paediatric exam.

Methods: 5th-year medical students (n=377) took a computer-based exam that included 6-9 KFPs and 9-20 crSBAs which assessed their clinical reasoning skills, in addition to an 11-station OSCE that assessed their clinical skills. Each KFP consisted of a case vignette and three key features using a 'long-menu' question format. We explored students' perceptions of the KFPs and crSBAs in 8 focus groups and analysed statistical data of 11 exams.

Results: Compared to crSBAs, KFPs were perceived as more realistic and difficult, providing a greater stimulus for the intense study of clinical reasoning, and were generally well accepted. The statistical analysis revealed no difference in difficulty, but KFPs resulted more reliable and efficient than crSBAs. The correlation between the two formats was high, while KFPs correlated more closely with the OSCE score.

Conclusions: KFPs with long-menu exams seem to bring about a positive educational effect without psychometric drawbacks.

INTRODUCTION

Various question formats have been described for the assessment of clinical reasoning (Higgs et al. 2008). The impact of each of these formats on student learning is still not well understood. Better understanding of this 'pre-assessment effect' (Cillier et al. 2012) would pave the way for clinical clerkship directors to better steer student learning through the concluding assessment. We therefore conducted a study comparing two different item formats, specifically key-feature problems (KFPs), consisting of a case vignette and three key features (KFs) using 'long menu' as response format, and context-rich single best answer questions (crSBAs) in terms of their impact on student learning and their relevant psychometric characteristics.

When comparing item formats several aspects are important. A test item essentially consists of two parts, that is, the stimulus and the response part (Schuwirth & van der Vleuten 2004). Whereas the former refers to the task imposed by the stem of an item, e.g. a case vignette, the latter denotes the method that examinees use to indicate their responses (Schuwirth & van der Vleuten 2004). The stimulus format can be either context-free or context-rich and shape the focus of the question (Schuwirth & van der Vleuten 2004). Context-free stimuli are almost exclusively aimed at measuring factual knowledge; context-rich stimuli, by contrast, serve to assess applied knowledge by presenting a specific scenario and asking for decisions, focusing on key features to solve a clinical problem, for example (as when a case vignette is used). The comparison we make in the present study is between two question formats with a *context-rich* stem designed to assess clinical reasoning.

As regards the response part, this can be grouped into two broadly defined categories: multiple choice-type questions (e.g. single best answer and multiple true/false questions) and open-ended (e.g. write-in) questions (Schuwirth & van der Vleuten 2004). Schuwirth et al. (1996b) demonstrated that electronic long-menu questions are an equivalent alternative to open-ended questions in computerised assessment. Long menus are alphabetically ordered long lists of (over 500) possible answers that prevent a cueing effect, because one has to type the solution into a dialogue field (Schuwirth et al. 1996b). The computer then searches through the long-menu list for 'hits'. The alternatives found are immediately presented to the examinee, so he or she can check whether the retrieved option is the desired one.

Besides varying in question types, context-rich items may also differ in the number of questions they contain: a case vignette for instance, may be followed by a set of multiple questions. One approach that is widely used to assess clinical reasoning, and that usually groups several questions together, is to use key features (Bordage et al., 1995; Page & Bordage, 1995). A key feature is defined as a critical step in the resolution of a problem, one where examinees are most likely to make a slip when trying to resolve the problem or one that complicates the identification and management of the problem in practice (Page & Bordage, 1995). Problems comprising a key feature, referred to as key-feature

problems (KFPs), consist of a brief stem with a short patient vignette (stimulus format) containing relevant and non-relevant elements, such as symptoms and findings, followed by one or more questions. KFP-based examinations, as such, do not follow a fixed item format, but should rather be seen as an *approach* to testing.

Studies on KFPs, however, have demonstrated that assessments yielded the best psychometric characteristics when responses were recorded in short-menu, write-in or electronic long-menu format and vignettes were followed by two to three questions each (Bordage et al. 1995; Page & Bordage 1995, Fischer et al. 2005, Norman et al. 2006, Hrynchak et al. 2014, Bronander et al. 2014). In a comprehensive review of the literature regarding the reliability and validity of KFPs, Hrynchak and colleagues (2014) conclude that published research supports the use of KFP-based examinations for the assessment of clinical reasoning. Their internal consistency reliability, as measured by Cronbach's alpha, has generally been reported to be acceptable. Face and content validity of the KF approach have been shown to be moderate to high, while construct validity has been shown to be good. Yet, the review acknowledges that all categories of validity evidence should be subjected to further scrutiny, as outcomes are contingent upon many contextual factors, including the population to which the test is administered, following which validity cannot be considered as a universal function of an assessment format. Considering the foregoing, and the fact that, to our knowledge, assessment at the *clerkship* level has received scant attention (Hatala & Norman 2002), we believe it important to delve deeper into this subject matter. Hatala and Norman (2002) introduced the KF assessment method into an *undergraduate* setting when evaluating the clinical decision-making skills of internal medicine clerkship students through a 2-hour paper-based exam consisting of 15 KFPs. They reported a Cronbach's Alpha of .49.

Another way to assess clinical reasoning is by means of context-rich single best answer (crSBA) multiple choice-type questions with case vignettes as stimulus format (Higgs et al. 2008). In such format the vignette usually is followed by only one question (Higgs et al. 2008), often containing 3 to 5 response options with only one correct answer (single best answer format, A-type questions).

At present, there is a paucity of evidence on the educational effects such different item formats have. It is widely anticipated that whatever item format is used, the stimulus format would be more important in determining what is tested than the response format (Schuwirth & van der Vleuten 2004). As single best answer question types are considered more beneficial in terms of their effectiveness, grading system, accountability and costs involved, these are often preferred to open-ended questions such as write-in or electronic long-menu questions (Elstein 1993, Downing 2002, 2009, Desjardins 2014). A few studies, however, have pointed out that the response format may also play a role, as cueing may influence item difficulty, with open-ended questions being more difficult than closed-ended questions (Hemskerk et al. 2008, Desjardins et al. 2014). These perceptions potentially influence the educational effect of an exam, making it interesting to investigate whether this is indeed the case, and if so, what this effect will be.

To our knowledge, no study has ever investigated the educational effect of using KFPs or crSBAs for the assessment of clinical reasoning in clinical clerkships. As it is generally accepted that assessment drives learning (McLachlan 2006, Cilliers et al. 2012, van der Vleuten & Schuwirth (2005)), we felt it would be valuable to investigate students' perceptions of these specific item formats which both use a context-rich stimulus but differ in terms of the response format (open-ended vs closed-ended) and to investigate whether these perceptions are also mirrored in the psychometric characteristics. Hence, the purpose of this study was to investigate students' perceptions of KFPs and crSBAs and compare both formats in terms of their impact on student learning and psychometric characteristics when used as a summative end-of-clinical-clerkship exam. Consistent with this aim we formulated the following research questions:

- a) How do students perceive KFPs with an electronic long-menu response format and crSBAs used for the summative assessment of their clinical clerkship, particularly when it comes to the (educational or pre-assessment) effect on their learning and how do they accept both formats?
- b) How difficult and reliable (*and* efficient) are KFPs compared to crSBAs? Do the results of these two formats correlate and how do these correlate with the total OSCE score in the respective end-of-clerkship exams?

METHODS

Context and Participants

Participants were fifth-year medical students (n=377) from Heidelberg Medical School doing regular paediatric clerkships, which included 10 Virtual Patients (VPs) that were integrated into the curriculum (Huwendiek et al. 2013) and specifically designed (Huwendiek et al. 2009) to foster clinical decision-making. At the end of their clerkship, students were subjected to a one-hour computer-based exam (Huwendiek et al. 2007), consisting of KFPs, context-rich and context-free SBAs. As context-free SBAs target factual knowledge rather than clinical reasoning (Schuwirth & van der Vleuten 2004), we did not include these questions in our study. Additionally, the end-of-clerkship assessment included an 11-station paediatric OSCE (5 min. per station, overall testing time also approximately 60 min.).

Instruments

We used the CAMPUS assessment software as assessment tool (Heid et al. 2006). This software tool facilitates the incorporation of many different question formats and has been used in the regular assessment of medical students for more than 15 years.

Key-feature Problems

Cases were constructed according to the key-feature approach (Page and Bordage 1995) and followed published recommendations (Farmer & Page 2005; Kopp et al. 2006). Each KFP consisted of a case vignette (context-rich stem) and three key features (KF) using 'long menu' as response format requiring free-text entry (Schuwirth et al. 1996b). In our study students selected an answer from the list by typing it into a dialogue box. The computer then searched through the long-menu list for 'hits'. The alternatives found were immediately presented to the examinee, so he or she could check whether the retrieved option was the desired one. An additional free-text comment field was available in case students did not find a suitable answer in the long menu. To encourage students to use the long menu, we informed them that they would earn only half of the points if they used the free-text comment field and that the long menu contained the correct answer.

Scoring Key for KFPs

We followed the recommendations of Page and Bordage (1995) and used a partial-credit score (a score between 0 and 1 that reflects the degree to which all correct responses were made). Page and Bordage (1995) found that reliability increased when partial-credit and summed scoring systems (the problem score is the sum of all question scores) instead of averaged (the problem score is the average of all subscores) or dichotomous (respondent receives 1 credit if he/she gives a predefined number of correct answers or 0 credits in all other cases) systems were used.

crSBAs

These questions consisted of a case vignette (context-rich stem) and focused on one relevant key feature. To answer each question, students had to select one correct answer from among a list of five options (A-type question).

For examples of the KFP and crSBA formats, see Appendix 1.

Blueprint

The Heidelberg catalogue of learning objectives (Bosse et al. 2011) served as blueprint for the assessment questions, with certain percentages of questions from relevant domains (e.g. paediatric infectious diseases, paediatric oncology and haematology, paediatric radiology, paediatric cardiology, general paediatrics, paediatric rheumatology, paediatric pulmonology) to ensure the exam covered a broad range of topics. KFPs mainly assessed eight important paediatric leading symptoms including 'fever without

focus', 'coughing', 'vomiting', 'diarrhoea', 'limping', 'changed urine', 'skin alteration' and 'oedema'. We specifically chose a KFP-based item format with long menu, as we hypothesised that the open response format would foster student learning more than a closed-question format would.

Review

Assessment questions were reviewed by at least two paediatricians and in many instances they were also subjected to scrutiny by a regional paediatric assessment alliance that regularly revised and exchanged those (Walter et al. 2008).

OSCE

The OSCE consisted of 10-11 paediatric stations assessing clinical skills (5 min. per station, overall testing time of 60 min.). The blueprint was comprised of three stations that focused on history-taking and communication skills with standardised patients playing the role of 'mothers' and 'fathers', three stations targeting the physical exam with mannequins, one to two media stations (e.g. characterising a neurological fit, facial paralysis), a percentile station, radiograph/sonographic interpretation station and procedural skills station (e.g. resuscitation, bladder puncture, spinal tap) (Bosse et al. 2006).

Focus Group Study

Three groups of approximately 40 clerkship students (n=116 in total) received an invitation to participate in a study on clinical education. Focus group interviews were held after students had completed both the rotation and the exam. To avoid bias, only those students who consented to participate received additional information about the study. Eventually, we randomly selected 39 students out of a total of the 65 students (56%) who had indicated their interest in the study and divided them into eight focus groups of approximately 4-7 students. Participants were offered a small compensation. During the interviews, we also explored other clerkship-related aspects the results of which have been reported elsewhere (Huwendiek et al. 2013).

Prior to analysis, the interviews were videotaped and transcribed verbatim. We selected a qualitative focus group methodology, as it has the potential to bring into focus not only the perceptions of participants but also the ideas and contemplations behind them (Krueger et al. 2000). Interviewing peers in groups rather than individually fosters a safe environment because it decreases the power distance between researcher and subjects and encourages in-depth discussion, which, in turn, may ultimately cause participants to change or adjust their views. To reduce any social pressure within the group that could prevent participants from speaking out freely, the moderator made clear that contributions of any kind would be highly valued and that each opinion

or view, no matter how divergent, would be respected. Prior to the discussion, participants were expected to write down their ideas about the theme under discussion. The lead author (SH) who is well versed in moderating focus group discussions conducted all the interviews which lasted two hours and took place on different days. Group interviews were guided by a questioning route (Appendix 2) which promoted consistency between interviews (Krueger and Casey, 2000). Two researchers (CD, FR) assisted the moderator by taking detailed notes and recording the interviews on video. We transcribed the tapes verbatim and sent participants the condensed scripts for approval.

Three authors (CD, FR, SH) subjected the scripts to a qualitative thematic analysis. This process consisted of a careful reading of the scripts, during which initial codes were spotted and singled out (Braun & Clarke 2006). From among these codes, themes and subthemes were identified and subsequently discussed in the research team in an iterative process, which elaborations served as new input for the coding process. These steps were repeated until all researchers were in agreement. Whenever possible, we kept an account of the strength of opinions and the number of times group members shared similar views. We ended the interviews after the eighth session, because the latter interview did not deliver us any new insights (Krueger & Casey 2000).

Statistical Analysis

We used the questions that assessed clinical reasoning (KFPs and crSBAs) of 11 end-of-clerkship summative exams (377 fifth-year medical students) as input for the statistical analysis. The exams all had the same blueprint but differed in the questions used. In line with recommendations by Page and Bordage (1995), we considered each case in the KFP component an item, and used a partial scoring system (a score between 0 and 1 reflecting the proportion of correct responses). We analysed difficulty (percentage of correct answers) at the item level for the pool of items over 11 exams and compared the difficulty distributions for the two question formats (KFPs and crSBAs). In order to investigate the reliability of the mean KFP score, we performed a generalisability analysis. For each exam we estimated variance components using a simple all-random person-by-item design. We then pooled variance components across exams weighted by sample size. The pooled variance components were used to estimate the generalisability coefficient (G coefficient), which is conceptually similar to Cronbach's alpha. The same procedure was applied to obtain the G coefficients of the crSBA questions. We were interested in comparing the reliabilities of KFPs and crSBAs that take equal amounts of testing time. To this end we also computed the G coefficients of KFPs and crSBAs for varying hypothetical amounts of testing time and computed efficiency using the Spearman-Brown prediction formula (Norman et al. 1996).

Test duration was determined by using the log files from the server data bank of the assessment system. To estimate how long students needed to answer a question, we measured the lapse of time from the moment students opened a page to a question

until they moved on to the next page (to the next question). Since KFPs consisted of three questions, we added the times of the respective three questions. If students went back to a specific question, we added all times they returned to this question. Times were added even when an answer remained unchanged.

We calculated Pearson's correlation for student's component scores for each of the 11 exams (exam level, student-level data). In order to reduce bias when estimating the average correlation, the scale of the correlations was Fisher-z transformed before calculating the average. To obtain true correlations we divided the observed correlations by the square root of the product of both reliabilities. Finally, we computed correlations of the two different written assessment formats with the total OSCE score. True correlations obtained after correction for attenuation should be interpreted with caution, because in the case of low reliability in particular, there may be a considerable risk of over-correction (Muchinsky 1996).

Ethical Approval

In Germany, where the present study was carried out, this type of educational study does not require approval from an ethics committee. Nevertheless, we confirm that participants took part on a voluntary basis, cannot be identified by the material presented and run no conceivable risk by having taken part in the study.

RESULTS

Focus Group Study

During the analysis, four main themes emerged, which are detailed below. Table 1 lists quotes from the interviews that are illustrative of each theme with the numbers in parentheses referring to the focus group and student number, respectively. Also see Figure 1 for a visual summary of the results.

Table 1: Quotes from focus groups:

Numbers in parentheses refer to focus group and student number, respectively

(i) KFPs were perceived as more realistic because of the long-menu questions*"KFPs are clearly more realistic, as you have to choose the correct answer yourself like in reality, and not from a predefined list". (1/2)**"In the KFP you must come up with an answer yourself, if you do not have one, you have a problem, which is also the case in real clinical practice". (5/6)**"If you are in an emergency situation, you do not have time to look up the correct therapy either, so I find these long menus more realistic than crSBAs". (3/1)***(ii) KFPs were perceived as more difficult to answer because of the long-menu questions***"As you have to know everything in the long-menu questions actively by heart, it is more difficult. The crSBAs are easier to solve as you will more easily recognise the correct answer from the options listed". (6/6)**"These long-menu questions were more difficult as you really had to reproduce your knowledge which is different from just having to recognise the correct answer". (5/4)**"When I had to give a diagnosis in a KFP, I found it much harder because I had to write it myself. In the crSBA I could see the options, which was easier". (4/1)**"When you enter your answer in the long menu it is harder to judge how well you have done in this question than in the crSBA questions". (1/4)***(iii) KFPs provided a greater stimulus for the intense study of clinical decision-making than did crSBAs***"As you knew that the 8 leading symptoms would be assessed with long-menu questions, you studied these leading symptoms much harder and more actively and thought about what they might ask you in the exam to be well prepared". (1/2)**"As I knew that 8 leading symptoms would be tested with long menu and I had to write it down myself in the exam, I studied these issues much more intensely, e.g. by working through the VP more carefully and repeatedly and taking notes while working through the cases". (7/4)**"I studied more for the long-menu questions. Had I known the exam was purely crSBA-based, I would have studied less, just the lecture notes, e.g., and I would have studied the main aspects of them, because this often suffices when answering multiple choice questions for which you only need to tick the correct answer". (5/5)**"Besides working through the VP several times I also explicitly wrote small guidelines in which I summarised for myself differential diagnoses that are typical of a leading symptom, typical diagnostic procedures and therapeutic options. Since I knew this would be relevant for the exam with KFPs I did that". (4/2)**"Long-menu questions in the VP or KFP much better show you your gaps in knowledge. This motivated me. In the crSBA it is easier to pretend that you know". (2/4)**"Learning with VPs was the best way to prepare for the KFPs, as you actively practised how to think and proceed in such a case. They taught me more than the problem-based learning tutorials did". (1/5)***(iv) Overall, KFPs were received positively and perceived to support the learning of clinically relevant topics; however, some aspects need to be taken into account when using them for high-stakes examinations.***"The KFPs were not so easy but I appreciate them better than crSBAs because they motivate you to learn relevant things which you will need in any case for your studies and future practice". (1/1)**"I liked the VPs and KFPs as I knew these would help me learn and test my knowledge of relevant topics, which is important to your future professional life, knowing how to proceed in cases of a patient with such a leading symptom". (4/4)**"I appreciated the KFPs as they showed you what you really knew and had understood and could apply and what not". (3/5)**"Sometimes I could not find the word I had in mind in the long menu - that was frustrating. At that point the additional free-text field was helpful". (1/4)**"I think it would be helpful if you could always give a comment on the item you have chosen from the long menu including why you think something is important. Similar to an oral exam...". (1/2)*

(i) KFPs were perceived as more realistic because of the long-menu questions.

Students perceived the KFPs with long-menu questions as more realistic than the crSBAs, as these required them to actively produce the solution as is the case in real life when there is no opportunity to choose from options.

(ii) KFPs were perceived as more difficult to answer because of the long-menu questions.

Students perceived the long-menu questions as more difficult to answer as there was no limited range of options to choose from. The long-menu response format required students to generate the answer themselves. Overall, students found the crSBA format to assess rather passive knowledge and the long-menu response format of KFPs to assess more active knowledge. Moreover, students had more difficulty judging how well they performed in the long-menu questions compared to the crSBAs.

(iii) KFPs provided a greater stimulus for the intense study of clinical decision-making than did crSBAs.

As the KFP with long-menu questions were perceived as more difficult and more realistic than crSBAs, students were more motivated to study hard and more in depth for the KFP assessment questions. Likewise, students reported KFPs motivated them more to use VPs for learning. More specifically, they worked through VPs several times and very ambitiously as they knew that these might be helpful in answering KFP questions. VPs were perceived as the best way to prepare for the KFPs with long-menu questions, whereas problem-based learning, seminars and learning from books were regarded as less helpful. Students found that actively working through the VPs (with embedded long-menu self-assessment options), including the helpful feedback they provided, prepared them well for the KFPs. Students generally felt that VPs afforded the best opportunities to actively train the clinical reasoning competencies required for solving the KFPs. Without these VPs, it would have been much harder to learn about clinical reasoning for the exam.

(iv) Overall, KFPs were received positively and perceived to support the learning of clinically relevant topics; however, some aspects need to be taken into account when using them for high-stakes examinations.

In general, KFPs were favourably received by students: They liked the use of KFPs for assessment purposes, for the incentive these provided them to learn important topics that were clinically and practically relevant. It is important that the content of the long

menus be exhaustive and include all synonyms relevant to each specific question. Omission of any synonyms caused frustration among students if they could not find the correct answer and had to produce synonyms themselves, a time-consuming effort in the exam which put additional pressure on them. The extra free-text entry field (as an escape option) was perceived as helpful when no solution could be found in the long menu.

However, students also voiced some concerns about KFPs. For instance, they regretted the fact they had to single out one correct answer phrased in only one or a few words without having the opportunity to elucidate their reflections, as would be the case in an oral exam. The inclusion of a free-text commentary field offered some solution, by allowing students to clarify their choices. Despite these concerns, however, students felt KFPs should remain part of the assessment, as they required knowledge of important and common leading symptoms, which, in turn, sparked intensive self-study. Yet, inclusion of KFPs was conditional on appropriate alignment of assessment contents with those of courses, including VPs.

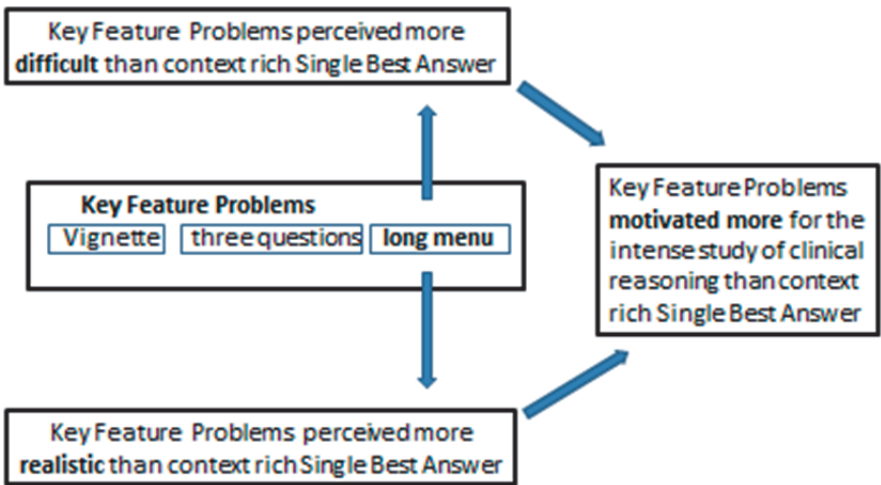


Figure 1: Visualisation of results of the focus group study. The arrows represent the influence. E.g., the long-menu response format of KFPs had an influence on the overall perceived realism and difficulty of KFPs.

Statistical Analysis

The results of the statistical analysis are presented in Tables 2-4. First, the analysis revealed that the two question types did not differ with respect to their level of difficulty (expressed as % of correct answers). Compared to crSBAs, the KFP questions presented the highest reliability, as assessed by a G coefficient in a D study. On average, the resolution of one KFP required 169 s, whereas 75 s were needed to solve one crSBA. Con-

sidering this, a test based exclusively on KFPs would require 45 min. of testing time to reach a .80 G coefficient, as opposed to 65 min. in a test that would be entirely crSBA-based. The correlation between the two formats was high, implying that both question types measure similar constructs (clinical reasoning). Table 3 presents the reliabilities to be expected for varying test durations, based on hypothetical tests composed of either one of the item types. Table 4 details the correlations of both written exam item types with the total OSCE score. KFPs correlated more closely with the total OSCE score than did crSBAs.

Table 2. Item- and exam-level characteristics

Item-level characteristics		
Items	KFP	crSBA
N items (all 11 exams)	81	162
Difficulty: Mean (Standard Deviation)	.77 (.13)	.75 (.20)
G analysis based on mean variance components over exams		
Items	KFP	crSBA
Mean number of items in the exams (rounded)	7	15
Reliability (G coefficient) for mean number of items in the exams	.65	.52
Number of items needed for a G coefficient of .80	16	52
Time needed to achieve a .80 G coefficient when using one question format only	45 min.	65 min.
Exam-level characteristics, N exams = 11		
Correlations	KFP - crSBA	
Correlation: Mean (Standard Deviation)	.50 (.11)	
Correlation after Fisher's Z-score transformation	.50	
True correlation after Fisher's Z-score transformation	.95	
Legend:		
KFP: Key-Feature Problem with Long Menu		
crSBA: context-rich Single Best Answer		

Table 3. Reliability per testing time (efficiency); average of 11 exams

Testing time	1 hour	2 hours	3 hours	4 hours
KFP	.84	.91	.94	.96
crSBA	.79	.88	.92	.94

Table 4. Correlations

	OSCE-KFP	OSCE-crSBA
Correlation M (SD)	.54 (.12)	.41 (.20)
Correlation after Fisher's Z-score transformation M	.55	.43
True correlation after Fisher's Z-score transformation M	.93	.81

DISCUSSION

In our search for an answer to the question of which assessments formats would be most suitable to test clinical reasoning in clerkship exams, we compared crSBAs and KFPs with respect to students' acceptance, their educational impact and psychometric characteristics in a mixed-methods study. From the analysis of the focus groups four themes emerged reflecting students' perceptions of KFPs compared to their crSBA counterparts: KFPs were perceived as more realistic and difficult, providing a greater stimulus for the intense study of clinical reasoning than did the crSBAs, and were generally well accepted, provided some preconditions were taken into account. The statistical analysis unveiled no difference in difficulty. KFPs exhibited a higher reliability and efficiency than did the crSBAs. The true correlation of the two written exam parts was high; however, KFPs correlated more closely with the overall OSCE score than did the crSBAs.

Our data indicate that KFPs with a long-menu response format provide a greater stimulus for the intense study of clinical reasoning than do crSBAs. This educational effect of KFPs has not been previously reported for clerkship students. A few studies, however, have demonstrated that the question format, notably the response format, can affect student learning (Cillier et al. 2012, Frederiksen 1984). Desjardins and colleagues (Desjardins et al. 2014), for instance, recently pointed out that the response format can indeed have an impact on the incidence of cueing, which again can impact perceived difficulty. In their study they made a comparison between multiple choice-type and open-ended questions, presenting a first group of students with open-ended questions, followed by multiple choice questions, and a second group with the same questions in the reverse order. Irrespective of the format seen first, multiple choice scores resulted higher than those of the open-ended questions. The observed pattern suggests that it was cueing rather than memory for prior questions that led to increased multiple choice questions scores. It may be hypothesised that if students truly believe that crSBAs are easier, an examination that is purely crSBA-based could be perceived as less challenging. This could impact the way students prepare for the examination, possibly studying at a more superficial level, which would surely oppose the intent of assessment, which should be to encourage a deep approach to learning (Al-Kadri et al. 2012, Newble & Jaeger 1983). That the response format matters is also supported by a think-aloud protocol study which found that more complex descriptions of thinking patterns are used when solving KFPs with long menu relative to SBAs (Schuwirth et al. 2001).

In terms of psychometric data, we have demonstrated that the measuring capacity of KFPs per unit of testing time is larger than that of crSBAs. With only 7 KFPs we reached a reliability (G coefficient) of .65, which means that we would achieve an estimated reliability of .80 with only 16 KFPs in a 45-minute exam. Hrynchak and colleagues (2014), in contrast, reported in their review on KFPs aimed at assessing clinical reasoning that internal consistency reliability of KFPs as measured by Cronbach's alpha would

be generally acceptable (i.e., between .70 and .95) if 25 to 40 KFPs were used (about 3-4 hours of testing time). In the case of undergraduate medical students, Fischer et al. (2005) found that a 90-minute, 15-KFP exam was able to achieve a reliability of .65 (Cronbach's alpha). Similarly, Hatala and Norman (2002) developed a 15-KFP 2-hour written exam to assess clinical decision-making skills in internal medicine clerkship students which exhibited an overall test reliability of .49. One explanation for the high reliability of the KFPs in the present study could be that KFPs always focused on eight important leading symptoms and were therefore restricted in terms of coverage. Another explanation could be that our KFPs were better aligned with the instruction process (Virtual Patients included). Their limited coverage could also explain why, psychometrically, KFPs were of the same difficulty as crSBAs, where the long-menu format is usually reported to be more difficult (Newble et al. 1979, Veloski et al. 1993).

The fact that KFPs correlated more closely with the total OSCE scores supports the results of the qualitative data. It may be hypothesised that KFPs with a long-menu response format assess active rather than passive knowledge, due to the absence of cueing just like in an OSCE, where there are no predefined answers to choose from either. Further studies are needed, however, to test this assumption. Using VPs for learning (formative assessment) and KFPs for the assessment of clinical reasoning is an example of constructive alignment of goals (to foster active clinical reasoning), methods that support active learning of clinical reasoning (VPs including long-menu questions for self-assessment) and assessment (KFPs which assess clinical reasoning 'actively' by long menu) (Biggs 1996).

One of the merits of the present study is that, to our knowledge, it is the first to demonstrate that KFPs with a long-menu response format offer clerkship students a powerful incentive to learn clinical reasoning. Other strengths are the fact that this study was performed in a real setting with many exams and students, and used a mixed-methods design. A limitation of this study is that focus group interviews are generally susceptible to bias, such as that due to the moderator's influence, although we took measures to minimise this. A further limitation is that the restriction of the coverage of the KFPs to eight leading symptoms may have impacted psychometric characteristics. We would therefore welcome replications of our study *without* such restrictions. Nevertheless, in terms of its educational value, our concept bore fruit in that students put a special emphasis on learning clinical reasoning for the most important leading symptoms as intended. Finally, we encourage studies focusing on one aspect only (either response format or stimulus format or number of questions per case vignette) to enhance our understanding even more.

CONCLUSIONS

Students perceived the KFPs with long-menu questions as providing a greater stimulus for the intense study of clinical reasoning than did crSBAs. Statistically, the KFPs revealed a higher efficiency than crSBAs. This study supports the idea that, from an educational perspective, both the stimulus and response format of questions seem to matter. Including KFPs with long menu in clerkship examinations seems to offer valuable opportunities to steer learning in clinical clerkships without psychometric drawbacks.

PRACTICE POINTS

- The analysis of focus group discussions revealed that students perceived KFPs with long-menu questions as providing a greater stimulus for the intense study of clinical reasoning than did crSBAs.
- Statistically, KFPs revealed a higher efficiency than crSBAs.
- This study supports the idea that, from an educational perspective, both the stimulus and response format of questions seem to matter.
- Including KFPs with long menu in clerkship examinations seems to offer valuable opportunities to steer learning in clinical clerkships without psychometric drawbacks.

ACKNOWLEDGEMENTS

We are grateful to our students who participated in the focus groups. We thank Jörn Heid (Heilbronn University, Heilbronn, Germany) for his support in calculating the times needed to answer questions and Daniel Stricker (Institute of Medical Education, Bern, Switzerland) for providing us with statistical advice.

We thank Angelique van den Heuvel for critically reading and correcting the English manuscript.

DECLARATION OF INTEREST

The authors report no conflict of interest. The authors alone are responsible for the content and writing of this article.

NOTES ON CONTRIBUTORS

Sören Huwendiek, MD, MME (Univ. Bern), is a Paediatrician, was one of the Curriculum Co-ordinators, Chairman of the Centre for Virtual Patients and E-learning Commissioner at Heidelberg Medical Faculty. He is now Head of the Department of Assessment and Evaluation at the Institute of Medical Education in Bern.

Friedrich Reichert, MD, is a Paediatrician at Olgahospital, Klinikum Stuttgart and wrote his MD Thesis about the design of Virtual Patients and their integration into medical curricula.

Cecilia Duncker, MD, is a Paediatric Resident at the University Hospital in Kiel and wrote her MD Thesis about the optimal integration of Virtual Patients into medical curricula.

Bas De Leng, MSc (Medicine) PhD (Medical Education), is an Educational Technologist and Chairman of the E-learning Competency Centre at the Institute of Medical Education, University of Münster, Germany.

Cees Pm Van Der Vleuten, PhD, is a Professor of Education, Scientific Director of the School of Health Professions Education (SHE) at Maastricht University, Maastricht, the Netherlands. He holds honorary appointments in the University of Copenhagen (Denmark), King Saud University (Riyadh) and Radboud University (Nijmegen).

Arno M.M. Muijtjens, MSc, PhD, statistician-methodologist, is an Associate Professor in the Department of Educational Development and Research, Faculty of Health, Medicine and Life Sciences, Maastricht University, the Netherlands.

Hans Martin Bosse, MD, MME, was responsible for alignment with the Paediatric clinical clerkship and is Specialist Registrar in the Department of General Paediatrics, Neonatology and Paediatric Cardiology, University Children's Hospital Düsseldorf, Germany.

Martin Haag, PhD, is a Professor of Software Engineering and Dean at the Faculty of Informatics at Heilbronn University, Germany. Furthermore, he is head of the Centre for Virtual Patients at the University of Heidelberg, Germany and speaker of the working group technology-enhanced learning of the German Society for Medical Informatics.

Georg F. Hoffmann, MD, is a Professor of Paediatrics and Chairman of the University Children's Hospital Heidelberg, Vice Dean of the Medical Faculty of the Ruprecht-Karls-University Heidelberg. He holds honorary appointments at the Medical Faculties of Padua, Italy, and Tongji, University Wuhan, China.

CHAPTER 6

Burkhard Tönshoff, MD, PhD, is a Professor of Paediatrics and Paediatric Nephrology, holds the position of a Vice Chairman of the Department of Paediatrics I, University Children's Hospital Heidelberg, Germany. He has been involved in the field of Virtual Patients for many years.

Diana H.J.M. Dolmans, PhD, is a Professor of Innovative Learning Arrangements within the School of Health Professions Education/Department of Educational Development and Research at Maastricht University, the Netherlands.

REFERENCES

- Al-Kadri HM, Al-Moamary MS, Roberts C, Van der Vleuten CP (2012): Exploring assessment factors contributing to students' study strategies: literature review. *Med Teach*. 2012;34 Suppl 1:S42-50. Review.
- Biggs J (1996): Enhancing teaching through constructive alignment. *Higher Education* 32: 347-364, 1996. Kluwer Academic Publishers. Printed in the Netherlands.
- Braun, V. and Clarke, V. (2006) Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3 (2). pp. 77-101. ISSN1478-0887
- Bronander KA, Lang VJ, Nixon LJ, Harrell HE, Kovach R, Hingle S, Berman N (2014): How we developed and piloted an electronic key features examination for the internal medicine clerkship based on a US national curriculum. *Med Teach*. 2014 Dec 11:1-6. [Epub ahead of print]
- Bordage G, Carretier H, Bertrand R, Page G. Comparing times and performances of French- and English-speaking candidates taking a national examination of clinical decision-making skills. *Acad Med* 1995;70:359-65.
- Bordage G, Brailovsky CA. Content validation of key features on a national examination of clinical decision-making skills. *Acad Med* 1995;70:276-81.
- Bosse HM, Dambe R, Juenger J, Kadmon M (2011): An interdisciplinary and interactive online tool to manage the continuous development of learning objectives in a curriculum. *Z Evid Fortbild Qual Gesundheitsw.* 2011;105(2):116-23. Review
- Bosse HM, Huwendiek S, Möltner A, Skelin S (2006): Making a pediatric OSCE fair and reliable. Congress of the Association for Medical Education in Europe (AMEE), Abstract 60 7, Abstractbooklet page 104, 2006
- Cilliers FJ, Schuwirth LW, van der Vleuten CP (2012): A model of the preassessment learning effects of assessment is operational in an undergraduate clinical context. *BMC Med Educ* 12:9.
- Desjardins I, Touchie C, Pugh D, Wood TJ, Humphrey-Murto S (2014): The impact of cueing on written examinations of clinical decision making: a case study. *Med Educ*. 2014;48:255-61.
- Downing SM. Assessment of knowledge in written test forms. In: Norman G, van der Vleuten C, Newble D, eds. *International Handbook of Research on Medical Education*. Dordrecht, Netherlands: Kluwer Academic 2002;642-72.
- Downing SM. Written tests: constructed response and selected response formats. In: Downing SM, Yudkovsky R, eds. *Assessment in Health Professions Education*. New York, NY: Routledge 2009;149-84.
- Durning S, Artino A, Schuwirth L (2013): Clarifying assumptions to enhance our understanding and assessment of clinical reasoning. *Acad Med* 2013; 88:442-8.
- Elstein AS. Beyond multiple-choice questions and essays: the need for a new way to assess clinical competence. *Acad Med* 1993;68 (4):244-9
- Farmer EA, Page G (2005): A practical guide to assessing clinical decision-making skills using the key features approach. *Med Educ*. 2005 Dec;39(12):1188-94.
- Fischer MR, Kopp V, Holzer M, Ruderich F, Juenger J (2005): A modified electronic key feature examination for undergraduate medical students: validation threats and opportunities. *Med Teach* 2005;27:450-5.
- Frederiksen N (1984): The real test bias: Influences of testing on teaching and learning. *Am Psychol* 39:193-202.
- Hatala R, Norman GR (2002): Adapting the key features examination for a clinical clerkship. *Med Educ* 2002; 36:160-5.
- Heid J, Bauch M, Brass K, Hess F, Jünger J, Haag M, Leven FJ (2006): Entwicklung und Einsatz eines sicheren Prüfungssystems für die medizinische Ausbildung. *GMS Med Inform Biom Epidemiol*. 2006;2(3):Doc10.
- Higgs J, Jones MA, Loftus S, Christensen N (2008): *Clinical Reasoning in the Health Professions*, 3rd edn. London: Elsevier (Butterworth Heinemann) 2008.
- Huwendiek S, Reichert F, Bosse HM, de Leng BA, van der Vleuten CPM, Haag M, Hoffmann GF, Tönshoff B (2009): Design principles for virtual patients: a focus group study among students. *Medical Education* 2009; 43: 580-588

- Huwendiek S, Duncker C, Reichert F, de Leng BA, Dolmans D, van der Vleuten CPM, Haag M, Hoffmann GF, Tönshoff B (2013): Learner preferences regarding integrating, sequencing and aligning virtual patients with other activities in the undergraduate medical curriculum: a focus group study. *Medical Teacher* 35:920-9
- Huwendiek S, Hanebeck B, Bosse HM, Haag M, Hoffmann GF, Tönshoff B (2007): Lernen und Prüfen mit virtuellen Patienten am Zentrum für Kinder- und Jugendmedizin des Universitätsklinikums Heidelberg: Ergebnisse der Evaluation im Rahmen des E-Learning-Preises Baden-Württemberg 2007. *GMS Med Inform Biom Epidemiol.* 2009;5(1):Doc10.
- Hrynchak P, Takahashi SG, Nayer M (2014): Key-feature questions for assessment of clinical reasoning: a literature review. *Med Educ.* 2014; 48(9):870-83.
- Kopp V, Möltner A, Fischer MR (2006): Key-Feature-Probleme zum Prüfen von prozeduralem Wissen: Ein Praxisleitfaden. *GMS Z Med Ausbild.* 2006;23(3):Doc50.
- Krueger RA, Casey MA. 2000. Focus groups: A practical guide for applied research. 3rd ed. London: Sage Publications, Inc.
- Muchinsky, P. M. (1996): The Correction for Attenuation. *Educational and Psychological Measurement*, 56(1), 63-75).
- Mayring P. 2010. Qualitative Inhaltsanalyse - Grundlagen und Techniken. 11. Beltz Verlag: Weinheim und Basel.
- McLachlan JC. The relationship between assessment and learning. *Med Educ* 2006;40:716–7.
- Newble DI, Jaeger K (1983): The effect of assessments and examinations on the learning of medical students. *Med Educ* 1983;17 (3):165–71.
- Newble DI, Baxter A, Elmslie RG. A comparison of multiple-choice tests and free response tests in examinations of clinical competence. *Med Educ* 1979; 13:263–8.
- Norcini J, Anderson B, Bollela V, Burch V, Costa MJ, Duvivier R, Galbraith R, Hays R, Kent A, Perrott V, Roberts T (2011): Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference. *Med Teach.* 2011;33(3):206-14.
- Norman GR, Swanson DB, Case SM. 1996. Conceptual and methodological issues in studies comparing assessment formats. *Teach Learn Med* 8:208–216.
- Norman G, Bordage G, Page G, Keane D (2006): How specific is case specificity? *Med Educ* 2006;40:618–23.
- Page GG, Bordage G (1995): The Medical Council of Canada's Key Features Project: a more valid written examination of clinical decision-making skills. *Acad Med* 1995;70:104–10.
- Page G, Broudo M, Blackmore D, Schulzer M, Bordage G (1992): Cueing as a factor In written examinations of clinical decision-making skills. In: Hart IR, Harden RM, Des Marchais J, eds. *Current Developments in Assessing Clinical Competence*. Montreal: CanHeal Publications 1992;184–91.
- Schuwirth L (2009): Is assessment of clinical reasoning still the Holy Grail? *Med Educ.* 2009;43(4):298-300.
- Schuwirth LWT, van der Vleuten CPM (2004): Different written assessment methods: what can be said about their strengths and weaknesses? *Med Educ.* 2004;38:974–9.
- Schuwirth L.W.T., Van der Vleuten, C.P.M., De Kock CA, Peperkamp, A.G.W., & Donkers H.H.L.M., (1996). Computerized case-based testing: a modern method to assess clinical decision making. *Medical Teacher* 18(4): 295–300.
- Schuwirth L, van der Vleuten CP, Donkers HH (1996a): A closer look at cueing effects in multiple-choice questions, *Medical Education*, 30, pp. 44–49.
- Schuwirth L, van der Vleuten CP, Stoffers HE, Peperkamp AG (1996b) Computerized long-menu questions as an alternative to open-ended questions in computerized assessment, *Medical Education*, 30, pp. 50–55.
- Schuwirth LWT, Verheggen MM, van der Vleuten CPM, Boshuizen HPA, Dinant GJ (2001): Do short cases elicit different thinking processes than factual knowledge questions do? *Med Educ* 2001;35:348–56.
- van Bruggen L, Manrique-van Woudenberg M, Spierenburg E, Vos J (2012): Preferred question types for computer-based assessment of clinical reasoning: a literature study. *Perspect. Med Educ* 2012;1:162–71.
- Van der Vleuten CP, Schuwirth LW, Driessen EW, Dijkstra J, Tigelaar D, Baartman LK, van Tartwijk J (2012): A model for programmatic assessment fit for purpose. *Med Teach.* 2012;34:205-14

- Van der Vleuten CPM (1996): The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ* 1996;1 (1):41–67.
- van der Vleuten CP, Schuwirth LW (2005): Assessing professional competence: from methods to programmes. *Med Educ*. 2005 Mar;39(3):309-17.
- Veloski JJ, Rabinowitz HK, Robeson MR (1993): A solution to the cueing effects of multiple choice questions: the Un-Q format. *Med Educ* 1993;27:371–5.
- Walter KN, Forster J, Scheerer U, Zieger B, Huwendiek S, Bosse HM (2008): Etablierung eines Prüfungsverbandes Pädiatrie in Baden-Württemberg. *Monatsschr Kinderheilkd*, 156, Suppl.1, 100, 2008.

APPENDIX 1

Example of KFP:

Vignette:

You are the physician in charge in a paediatric outpatient department. A parent presents the 4-year-old girl Jessica who refuses to walk. She has been having reoccurring upper respiratory tract infections for 4 weeks. In the physical exam she appears pale.

Although she refuses to walk, muscular strength, neurologic and joint findings of the lower extremities are normal. All other physical exams are equally normal.

Question 1:

After having taken a full history and having performed a physical exam:

What would be your next step in the investigation? (Please provide one answer; be as specific as possible)

Question 2:

In the blood count including differentiation you find the following: leucocyte count 2.8/nl (4,5–13/nl); erythrocyte count 2.0/pl (3.9–5.3/pl); haemoglobin 6.1 g/dL (11–14.5 g/dL); haematocrit 18% (31–37%); thrombocyte count 90/nL (180–530 /nL). The differentiation of the blood count is still pending.

What is the most likely diagnosis in this case?

Question 3:

How do you confirm your suspected diagnosis?

Example of crSBA

You are a paediatrician in your private practice. A parent presents a three-year-old girl who has suffered from dizziness, nausea and vomiting since 2 weeks. The physical exam reveals a gait ataxia and postural instability.

What should be done first?

- physiotherapy
- referral to a child and youth psychiatrist
- referral to an ophthalmologist and ENT specialist
- Lumbar puncture
- MRT of the skull

APPENDIX 2

Questioning route

1. How did you perceive the electronic clerkship exam?
2. How did you perceive the different assessment formats in the clerkship exam?
3. Did the different assessment formats have an impact on your learning? If yes, how?

CHAPTER 7

General discussion

Clinical reasoning is generally considered a decisive component of nearly everything doctors do in practice (Higgs et al. 2008, Norman 2005, Eva 2005). However, medical education seems to fall short of providing students with adequate opportunities to learn these skills, exposing them to too few and less varied real patients (Rattner et al. 2001) and of providing adequate feedback and coaching (Wimmers et al. 2005, Schmidt & Mamede 2015, Eva 2004). Besides using real patients, one may also resort to the use of Virtual Patients, which show promise as a tool to foster the learning of clinical reasoning (Cook et al. 2010). However, there is a paucity of concrete evidence on how to design and implement VPs with the aim to foster clinical reasoning (Cook et al. 2010). Therefore, we investigated the following two main research questions in this dissertation:

I: How to improve the design of Virtual Patients to foster learning, especially of clinical reasoning?

II: How to improve the implementation of Virtual Patients to foster learning, especially of clinical reasoning?

In this chapter we first discuss the main research findings and conclusions of this dissertation and then consider its strengths and limitations. Finally, we discuss the implications for further research and for practice.

DISCUSSION OF MAIN RESEARCH FINDINGS AND CONCLUSIONS

First we discuss the main research findings of Chapters 2-4 in order to answer the first main research question, and then those of Chapters 5-6 to answer the second main research question.

Main research question I: How to improve the design of VPs to foster learning, especially of clinical reasoning?

Based on Chapters 2-4 we conclude that the following three main aspects seem to be especially relevant in the optimisation of VP design to foster the learning of clinical reasoning: the use of instructional design criteria, the incorporation of virtual guidance on clinical reasoning and the provision of validated instruments for systematic further improvements of VP design. The first two aspects have been addressed, with differing degrees of detail, by different stakeholders, students and VP experts in Chapters 2-4. The third aspect has been addressed in Chapters 3 and 4 building also on results of the previous chapters. In the following we elaborate on these main aspects.

Using instructional design criteria

Main criteria found in the studies of Chapters 2-4 were that VPs should be relevant, of appropriate difficulty, interactive, provide specific feedback, make appropriate use of different media, help students focus on relevant learning points, and be authentic in terms of student tasks. Further, it has been demonstrated in Chapter 2 that students, while appreciating a mix of questions to shore up interactivity, prefer Long-Menu to e.g. Short-Menu questions in a VP, because they find these questions more challenging, more suitable for the identification of gaps in their prior knowledge, and more motivating as they have to generate their response by themselves.

Most of these findings reflect criteria from outside the VP design literature, such as principles for the development of teaching cases ('interactive', Kim et al. 2006), lessons from simulation research ('provide specific feedback', Issenberg et al. 2005) and principles for multimedia learning design including cognitive load theory ('help students focus on relevant learning points', Grunwald & Corsbie-Massay 2006, Stemler 1997, Mayer 2005, Lawless 1997, Merriënboer & Sweller 2010). For details, refer to Table 1, Chapter 1, page 16.

To our knowledge, however, no study has ever compared the effects of Short- and Long-Menu questions on learning with VPs. Chapter 2 reveals that students prefer Long-Menu questions for the reasons outlined above. This resonates with results from assessment research, where open response questions like LMQs are perceived as more difficult and motivating (Desjardins et al. 2014, Frederiksen 1984, see also Chapter 6).

Some of the criteria found also address conflicting goals. To name an example, the students of Chapter 2 expressed the wish to make all decisions a doctor would make in real life (authentic tasks), while at the same time they appreciated assistance to help them 'focus on relevant learning points'. Finding the right balance here is surely challenging and depending on students' prior knowledge and their stage in a curriculum as reflected in the 'appropriate difficulty' criterion and other studies from clinical reasoning research (Durning et al. 2012). This conundrum, bearing a close relation to clinical reasoning, leads us to the next main aspect of VP design on how to incorporate virtual guidance on clinical reasoning into a VP.

Incorporating virtual guidance on clinical reasoning

The incorporation of clinical reasoning support into the VP, by asking explicit questions and by guiding the clinical reasoning process, was perceived by advanced students as very helpful in that it allowed the learning of clinical reasoning, as evidenced by the focus group study of Chapter 2. Such coaching would take the form of a virtual supervisor asking the respective questions, such as 'generate a differential diagnosis' early in the case, eliciting discriminating features of a set of diagnostic hypotheses, asking for an early synthesis, and providing students with feedback on their answers. Students indi-

cated that this coaching, together with the instructional design principles of ‘interactivity’, ‘specific feedback’ and ‘authenticity of student tasks’, made them feel well prepared for dealing with real patients to an extent no other type of teaching had previously achieved.

Incorporating virtual guidance seems foundational to the design of VPs, but it is also the most laborious part. Such construction could only thrive on a good collaboration between educational and medical experts in the field. The above-mentioned conclusions resonate well with strategies to foster clinical reasoning in advanced students from outside of the VP literature as described in Table 1 in Chapter 1 (page 16, Bowen 2006, Nendaz & Bordage 2002; Bordage 1999, Schmidt & Mamede 2015) and a recent ‘twelve tips publication’ on VP design (Posel et al. 2015). In their recent review and proposal on teaching clinical reasoning, Schmidt & Mamede (2015) propose that advanced students need to see many patients to learn the full range of different manifestations of each disease and those of related diseases showing similar symptoms and to be able to recognise and distinguish between them. They call for a ‘deliberate reflection strategy’, by which students actively weigh various diseases and hypotheses against the signs and symptoms presented in a patient’s history (Schmidt & Mamede 2015). This deliberate reflection strategy is exactly what we describe as ‘virtual guidance on clinical reasoning’ within a VP and what students perceived as extremely helpful in Chapter 2. It was recently reported that this approach, albeit tested outside a VP context, leads to better clinical reasoning in future diagnostic probes (Mamede et al. 2014, Schmidt & Mamede 2015).

By implementing strategies that were originally developed for other settings, transposing them to VPs and evaluating how these were perceived by students, we could add to the existing body of VP design literature.

Providing validated instruments for systematic further improvements of VP design

To systematically improve VP design, well-designed instruments can be helpful. In Chapters 3 and 4 respectively we developed such instruments, including a VP design typology and a VP design survey tool.

To develop the **VP design typology** reported in Chapter 3, we synthesised 19 factors relevant to VP design around four categories based on the literature, a modified Delphi study by international VP experts and a validation process. The category ‘general’ included topics such as ‘title’ and ‘typical study time’, the category ‘educational’ topics such as ‘educational level’ and ‘objectives’, the category ‘instructional design’ topics such as ‘interactivity use’ and ‘feedback use’ and the category ‘technical’ topics such as ‘originating system’ and ‘dependence’.

This typology can help VP developers to gain more clarity about relevant VP design features and can direct researchers in selecting those features that deserve reporting

when publishing VP-related research. If the scientific literature does not exactly state how the intervention (e.g. a VP) was designed, it is very hard to draw the right conclusions from such research (Cook et al. 2007). For example, the type and number of questions included can vary significantly between different VPs and VP systems. It is therefore important that authors report the type and number of questions included in a VP, for which they may use the described typology ('instructional design' category, 'interactivity use' factor) to allow for correct interpretations by readers and in possible future systematic reviews. However, reporting deficiencies not only seem to pervade the VP literature: According to a systematic review by Cook et al. (2011), the quality of reporting scientific results seems to be generally critical in medical education research. While there are several guidelines for the reporting on qualitative research, for example, (O'Brien et al. 2014) no guidelines exist as yet for the reporting on e-learning interventions or VPs. Considering that the VP research field is still in its infancy, it seems especially important to have such guidance in the form of a typology. Since its publication in 2009 the VP typology has been cited in upwards of 65 publications, which confirms the need for such an instrument. We are convinced that this typology will help move this research field beyond its infancy.

To develop the **VP design evaluation instrument** documented in Chapter 4, we availed ourselves of the published evidence including the lessons learned from Chapters 2 and 3 on VP design. The short instrument that resulted consists of six questions distributed between three factors (authenticity of patient encounter and the consultation, cognitive strategies in the consultation, coaching during consultation) and one global score. It is theory-based and has been demonstrated to validly measure students' perceptions of VP design, especially in relation to its purpose to foster clinical reasoning, as long as the sample used per VP is large enough. Although the evaluation of one VP requires at least 200 student ratings before the outcome for all three factors can be considered reliable, the instrument is still unique in its kind (Huwendiek et al. 2014). Since the same VPs are increasingly used by large numbers of students (Berman et al. 2008, 2011, Poulton et al. 2007; Balasubramaniam et al. 2009; Poulton & Balasubramaniam 2011), it seems feasible to achieve the required number of evaluations. If we wield this short VP design evaluation instrument to collect feedback from many users alongside the previously described typology (Chapter 3) to characterise the VP we wish to evaluate, we may be able to secure systematic VP design improvements in the future, for VP design and perception can be easily combined in order to make the right inferences.

Main research question II: How to improve the implementation of Virtual Patients to foster learning, especially of clinical reasoning?

Based on Chapters 5-6 on VP implementation, we conclude that the following three main aspects are especially relevant to VP implementation that aims to foster the learn-

ing of clinical reasoning: Sequencing according to complexity, aligning instruction and aligning assessment. In the following we elaborate on these main aspects.

Sequencing according to complexity

In Chapter 5 we found that students perceived it helpful when they were exposed to VPs with increasing complexity, adapted to their knowledge. They also deemed the following order of VPs and other educational activities appropriate to maximise their learning of clinical reasoning: i) acquire a knowledge base by means of a symptom-oriented lecture or computer-assisted learning activity, ii) work on 1 to 2 VP(s) to apply the newly gained knowledge, iii) in the case of complex diseases, hold a tutor-led small group discussion, and iv) meet a real patient presenting the same leading symptom.

The reasons students gave for their sequence preferences appear to be in line with instructional design principles, such as ‘from simple to complex’ and ‘from low to high fidelity’ (Van Merriënboer & Sweller 2010) and with the results of meta-analyses of simulation research that demonstrated the importance of steadily increasing complexity adapted to learners’ abilities (Issenberg et al. 2005; McGaghie et al. 2010).

Aligning instruction

Students also stressed the importance of aligning VPs and other educational formats to foster clinical reasoning, as reported in Chapter 5. VPs and related educational activities should use appropriate instructional methods and address aspects to which they are particularly suited. Strategies to foster clinical reasoning (Bowen 2006, Schmidt & Mamede 2015) should be implemented both in VPs (Chapter 2) and in related educational activities (Chapter 5). In a small face-to-face group session, for example, instead of revisiting the preceding VP(s) tutors should challenge students to discuss the differential diagnosis including confirming and disconfirming factors and feedback. In bedside teaching students should take the history and perform the physical examination, while supervised by clinical teachers who challenge them with questions about clinical reasoning. To fulfil their job, teachers should be properly trained for their tasks. This echoes the work by Eva (2004) who concludes that clinical teachers should not rely on students to make meaningful comparisons across problems spontaneously. Successful reasoning by way of analogy is much more likely to occur when students have been explicitly instructed by VPs and clinical supervisors to attempt to identify similarities in the underlying concepts of superficially distinct problems (Thompson et al. 2000).

Overall, this resonates well with the idea of a flipped classroom approach (Prober & Heath 2012, Prober & Khan 2013) in that it uses the precious face-to-face time for learning activities which can only be done face-to-face (e.g., seeing a real patient in bedside teaching) and reserves other educational activities (e.g., online VPs) for self-study. Similarly, it is consonant with the literature on the alignment of computer-

mediated activities with other teaching activities (Bunk & Graham 2006, Choules 2007, Bonk 2006). Sequencing and aligning instruction (as outlined above) were considered to be especially helpful in fostering clinical reasoning and preparing students optimally for clinical reasoning in real patients.

Aligning Assessment

Students recommended that VP-related content be reflected in the assessment (Chapter 5). Not only is this consistent with general learning and assessment principles (Larsen et al. 2008; Schuwirth et al. 2011) and a study on the integration of VPs (Fischer et al. 2008), it also demonstrates that students recognise and appreciate the practical relevance of VPs. They also judged the 'Key-Feature Problems with Long Menu' response format used in a summative clerkship exam a greater stimulus for the intense learning of clinical reasoning (with VPs) than the context-rich Single Best Answer questions (Chapter 6). This educational (pre-assessment) effect of KFPs has not been reported earlier for clerkship students. However, some studies do suggest that the response format can have an influence on students' learning (Cillier et al. 2012, Frederiksen 1984). Desjardins and colleagues (2014) recently demonstrated that the response format can indeed have an impact on whether cueing occurs, which, in turn, can affect perceived difficulty. Including KFPs with Long Menu in summative clerkship exams seems to bring about educational effects without psychometric drawbacks.

The design and implementation of VPs for active learning of clinical reasoning (Chapters 2 & 5) and use of KFPs with Long Menu for the assessment of clinical reasoning (Chapter 6) described in this dissertation is also an example of constructive alignment, specifically of the goals (to foster active clinical reasoning), the methods (VPs and implementation that support active learning of clinical reasoning) and the assessment (KFPs that 'actively' assess clinical reasoning using Long Menu) (Biggs 1996). This alignment is said to optimally support the intended learning (Biggs 1996).

STRENGTHS AND LIMITATIONS

Major strengths of this dissertation are the authentic settings in which studies were performed, the involvement of different stakeholders in the studies and the combination of different research methodologies. In the following we will elaborate on these aspects.

Three of the five studies (Chapters 2, 5, 6) were conducted in an authentic educational practice setting (specifically a paediatric clinical clerkship), which enhances the value of findings for educational practice. Moreover, students were exposed successively to VPs and real patients and could therefore make better judgements than would have been the case if the studies had not included real patients. After each study, we

used the results to continually improve the way VPs were used in this clinical clerkship. As such, this dissertation in effect represents a kind of design-based research (Anderson & Shattuck 2012).

Different stakeholders were involved. While in most studies students were the target of our investigation, in Chapters 3 and 4 VP experts and medical teachers, national and international, were strongly involved in the development of both the VP typology and the VP evaluation instrument. Exploring different stakeholders' perceptions of the same aspect, e.g. VP design, can improve the validity of the results.

Furthermore, we used different research methods in this dissertation. The results of the focus group study on the design of VPs (Chapter 2) also informed the Delphi-study among VP experts to develop the VP typology (Chapter 3) and the questionnaire study on VP design (Chapter 4). In Chapter 4 we used data from VP experts, think-aloud protocols from students and statistical measures to establish validity evidence. In Chapter 6 we used qualitative data from a focus group study and quantitative psychometric data resulting in a mixed-methods design similar to that of Chapter 4. By blending methods we were able to achieve a more comprehensive picture and enhance our understanding of the issues under scrutiny.

Major limitations of this dissertation are that three studies were conducted at only one centre, that it did not include a study on the long-term implementation of VPs, or one on the transfer of learning to real patients using objective measures. In the following we will elaborate on these aspects.

Three studies (Chapters 2, 5, 6) of this dissertation emanate from senior undergraduate medical students (clerkship students) in one specialty – Paediatrics – from the same medical school. Given the possible uniqueness of the curriculum of this medical school and its students, it is not inconceivable that medical students from this context differ from their counterparts in other parts of the country or even the world. Generalisation to more junior students, to students in other specialties and students from other countries is therefore questionable.

The studies of this dissertation focused on learning at one point in time (Chapters 2, 5, 6). We did not investigate the longitudinal integration of VPs e.g. throughout a whole curriculum. As longitudinal repetitive practice is central as evidenced by the principles of 'deliberate practice' (McCaghie et al. 2011) and 'mastery learning' (Wayne et al. 2006), this should also be done with VPs to foster learning of clinical reasoning. Further, students at different levels of training may need different ways of clinical reasoning teaching (Schmidt & Mamede 2015), which calls for different VP designs to optimise their learning.

Finally, we did not objectively measure the transfer of learning to practice with real patients. While students indicated that this would occur in the studies in Chapters 2, 5 and 6 and we also did some objective measurement of learning from VPs in Chapter 6, we did not objectively measure the transfer of learning to practice with real patients.

IMPLICATIONS FOR FUTURE RESEARCH

Suggestions for future research can be derived from our considerations in the limitations section. Future research into VP implementation should be extended to other contexts, carried out in multiple centres and over longer periods of time, and investigate the transfer of learning to practice with real patients using objective measures. In the following we will elaborate on these aspects.

In order to discover if the results of this dissertation can be extrapolated to other settings, the following two studies would be interesting. First, a multicentre study consisting of the following three phases: i) development and design of a set of VPs along the lines specified by Chapter 2 and based on the content of a respective national core curriculum (e.g., Paediatrics, Internal Medicine); ii) implementation and assessment of VP learning in accordance with the results of Chapters 5 and 6; and iii) evaluation of VP design and implementation of the previous phases using different methodologies, including the validated instrument of Chapter 4. Such a study in the field of Paediatrics is currently in preparation based on the Paediatric core content established during a Delphi study in Germany (Huwendiek et al. 2011). To increase the chances of success, lessons learned from a similar project in North America should also be taken into account (Berman et al. 2008 & 2011, Schifferdecker et al. 2012). Second, it could be investigated whether the lessons learned from the use of VPs in undergraduate education could also be extrapolated to postgraduate education and perhaps even to other professions. One example currently under way is the interprofessional paediatric emergency training at the Children's Hospital of Heidelberg. This training for physicians and nurses consists of three parts: (i) provision of paper-based paediatric emergency guidelines, (ii) work up of 8 VPs in self-study, and (iii) interprofessional practical simulation team training. First results of this project indicate that the lessons learned from this dissertation can generally be extrapolated (Huwendiek et al. 2012) and that patient treatment can be improved through this training programme.

Principles of 'deliberate practice' (McGaghie et al. 2011) and 'mastery learning' (Wayne et al. 2006) implicate that one has to practise purposefully again and again with feedback to acquire expertise and become proficient e.g. in 'clinical reasoning'. Therefore, long-term implementation of VPs seems promising. When intending to deploy VPs for a longer period of time to foster clinical reasoning, the following aspects could guide implementation: VPs should be designed to foster clinical reasoning (e.g. Chapter 2). However, in a recent review Schmidt & Mamede (2015) suggest that the teaching of clinical reasoning should be best tailored to learners' knowledge and expertise. More specifically, for beginners the focus should rather be on developing causal knowledge, explaining disease in terms of pathophysiological principles, for intermediates on the encapsulation of pathophysiological knowledge, whereas for more seasoned students, such as clerkship students, it should be on the development of illness scripts, e.g. by using the strategies described in Chapter 2 (Schmidt & Mamede 2015). Additionally, VP

implementation strategies should be taken into account (e.g. Chapter 5) and Key-Feature Problems using Long Menu as assessment method should be integrated longitudinally (e.g. Chapter 6). Furthermore, it has been repeatedly demonstrated that learning is more effective when students are exposed to a host of different problems at once ('mixed practice') than when problems are administered on a (diagnostic) category by category basis (Eva 2004). This should also be taken into account when implementing VPs longitudinally. VP design and implementation could be further improved by using the design questionnaire described in Chapter 4 and the curriculum integration questionnaire reported elsewhere (Huwendiek & de Leng 2010); however the latter should first be formally validated. This curriculum with longitudinally implemented VPs could then be compared with a curriculum without VPs that is otherwise similar, e.g. using a clinical reasoning progress test (Williams et al. 2011). Of course, many possibly biasing factors would then need to be taken into account, which would make such a comparison challenging. One example of such a longitudinal VP implementation is the longitudinal VP project at Heidelberg Medical Faculty (see Figure 3, page 133). In this project most of the above-mentioned features have been taken into account. It deploys over 100 VPs in 11 different subject areas in a learning spiral to foster clinical reasoning. While in the first years the focus is more on biomedical and pathophysiological knowledge (e.g. glucose regulation) within a VP presenting typical features (e.g. of diabetes mellitus), later the emphasis shifts towards coaching of clinical reasoning, for instance by contrasting features of different cases to foster the development of illness scripts (see Chapter 2). First results confirm the value of this kind of implementation; however the impact of this overall project has not yet been evaluated.

Another question worth exploring using objective measures is whether the use of VPs can improve patient care. This task can be quite daunting. However it may be possible to answer this question, for instance by drawing a comparison between patient outcomes in a clinic where doctors were trained in clinical reasoning the classical way, e.g. by consultant supervision, seminars and workplace-based formative assessments (MiniCEX) (Kogan et al. 2009), and those in a clinic with comparable patients where doctors were additionally trained by VPs designed and implemented in accordance with the conclusions of this dissertation. Such a challenging endeavour would require that many aspects be controlled for, like severity of illness, physician characteristics, clinic characteristics and overall supervision and training time. However, it should be possible. Norcini et al. 2014 for example demonstrated that, after adjustment for several relevant factors, performance on the Step 2 Clinical Knowledge exam had a statistically significant inverse relationship with patient mortality. Each additional point on the exam was associated with a 0.2% (95% CI: 0.1%-0.4%) decrease in mortality. While this study investigated the association between exam results and patient mortality, a similar attempt could be made by exploring the association between different training formats (with/without VPs) and patient outcomes.

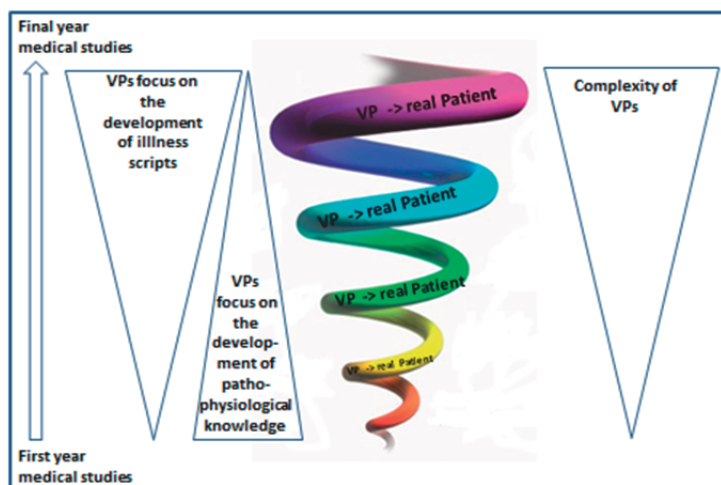


Figure 3: Longitudinal Virtual Patient Project at Heidelberg to foster clinical reasoning

The project integrates upwards of 100 Virtual Patients in the first year through to the final year. Displayed vertically in the figure, VPs' focus on fostering the development of illness scripts and the complexity they offer steadily increase from year one of medical studies to the last year. In a similar fashion, the focus on patho-physiological knowledge decreases. This approach embraces Schmidt and Mamede (2015) and Norman's (2009) suggestion that such a teaching strategy would support the transfer of knowledge from the first to later years and the development of clinical reasoning. Displayed rather horizontally in the spiral, representing clerkships or a study year, for instance, VPs are also integrated with increasing complexity and used to prepare students for the more complex real patients (Chapter 5).

IMPLICATIONS FOR PRACTICE

The main implications for medical education appear to be that when deploying VPs, they should be well designed, systematically further improved and well implemented to optimally foster clinical reasoning. In the following we will elaborate on these aspects.

VPs should be well designed, meeting instructional design criteria that ensure, for example, an appropriate level of difficulty, authenticity, interactivity, feedback and a focus on relevant learning points (Chapter 2). Moreover, virtual guidance on clinical reasoning should be incorporated into VPs in the form of questions asking for discriminating and confirming features to support learning of clinical reasoning (Chapter 2). As the development of well-designed VPs is laborious it should be considered to develop them collaboratively and multicentre as outlined in the 'implications for future research' section of this Chapter. Especially laborious is the development of VPs' content and educational design, not the technical design, as there are ready-made software programmes available. It is therefore recommended to recruit sufficient manpower for the content and educational design.

When conducting research and reporting on studies involving VPs it is advisable to specify the general aspects of these VPs, such as their typical study time and providing a short description, their educational aspects, such as educational level and objectives, their instructional aspects, such as interactivity and feedback use, and their technical aspects, such as originating system and dependence. In doing so, one could use the described typology to support the quality of reporting and the scientific progress concerning VP design (Chapter 3). The above-mentioned aspects can be easily addressed in an appendix to a publication, in which all VPs used are characterised. When having your VP design evaluated by medical students, use a short, valid and reliable instrument that focuses e.g. on the authenticity of the encounter, cognitive strategies and coaching during the encounter to allow for systematic further improvements of VP design through learner feedback (Chapter 4). However, in settings where less than 200 judgments per VP are to be expected, other ways to collect feedback -e.g. qualitative- might be more appropriate to make the right conclusions.

VPs should be well implemented, meaning that they should be sequenced appropriately, aligned with other educational activities including bedside teaching with real patients and with teacher training to foster learning of clinical reasoning (Chapter 5). Long-term implementation should also be considered to allow for expertise development and sustainable learning as outlined in the 'implications for future research' section of this chapter. Teaching and assessment of learning with VPs should be well aligned by using methods that assess clinical reasoning rather "actively", such as a free-text or Long-Menu response format to capitalise on the educational pre-assessment effect of these formats (Chapter 6). The Long-Menu response format has the advantage over the free-text format in that it allows for a computerised analysis of candidates' answers necessitating only some cross-checking e.g. to detect possible other correct answers.

It is important to keep in mind that VPs should never be a replacement but an add-on to the learning of clinical reasoning through real patients supported by medical teachers and clinical supervisors (Chapters 2 & 5). Whether there are VPs available in a setting or not, it would in any case be great if medical teachers and clinical supervisors could take the present recommendations on how to foster learning of clinical reasoning into account during face-to-face sessions with students and trainees (Bowen 2006, Schmidt & Mamede 2015).

Summary

Chapter 1 forms the introductory chapter of this dissertation, in which we describe its scientific background, rationale, aim and overarching research questions.

Clinical reasoning enables doctors to establish the right diagnosis and make the right treatment decisions for a patient. Clinical reasoning is generally considered a decisive component of nearly everything doctors do in practice. The clinical reasoning skills of a physician have as such a great importance and relevance to patient outcomes and to patient safety. There are indications that clinical reasoning is not well taught at present. Literature shows that so-called Virtual Patients (VPs) are especially suited to foster learning of clinical reasoning. Virtual Patients are online cases in which the learner takes on the role of physician, has to make all decisions him/herself and receives feedback on these decisions. They have been defined as ‘interactive computer simulation[s] of real-life clinical scenarios for the purpose of medical training, education, or assessment’. When learning with electronic learning tools such as VPs, two aspects have shown to be of great importance: their design and implementation.

When we embarked on this research, there was a paucity of literature on how to design and implement Virtual Patients to foster clinical reasoning. That is why we began our research project aimed at providing evidence for how learning with Virtual Patients to foster clinical reasoning in medical students can be improved by addressing the following two main research questions:

I: How to improve the design of Virtual Patients to foster learning, especially of clinical reasoning?

II: How to improve the implementation of Virtual Patients to foster learning, especially of clinical reasoning?

In **Chapter 2** we explore what students perceive as the ideal features of Virtual Patient design to foster their learning with a special focus on clinical reasoning. After exposure to at least eight Virtual Patients varying in design, students discussed Virtual Patient design in focus groups. According to the students, learning is facilitated when Virtual Patients are relevant, of appropriate difficulty, interactive, provide specific feedback, make appropriate use of different media, help them focus on relevant learning points, foster recapitulation of key learning points, are authentic in terms of the web-based interface and student tasks and ask questions and offer explanations that enhance clinical reasoning. After working through well-designed Virtual Patients, students felt very well prepared for reasoning in real patients, better prepared than with any other existing method they experienced before. We conclude that students perceived the design principles identified as being conducive to their learning and that many of these principles are supported by the results of published studies outside of VP research and that further studies are needed.

In **Chapter 3** we describe the development of a Virtual Patient typology based on the literature, a review of existing VP systems and an international consensus process among Virtual Patient experts using a modified Delphi technique. In this empirically derived Virtual Patient design typology, we synthesised 19 factors around 4 categories deemed relevant to Virtual Patient design: general (title, description, language, identifier, provenance, typical study time); educational (educational level, educational modes, coverage, objectives); instructional design (path type, user modality, media use, narrative use, interactivity use, feedback use); technical (originating system, format, integration and dependence). We conclude that this empirically derived Virtual Patient design typology provides a common reference point for all those wishing to report on or study Virtual Patients.

In **Chapter 4** we report on the design and validation of a Virtual Patient design evaluation instrument taking the published evidence including the lessons learned from Chapters 2 and 3 on Virtual Patient design to foster learning of clinical reasoning into account. The established short instrument consists of six questions distributed between three factors (authenticity of patient encounter and the consultation, cognitive strategies in the consultation, coaching during consultation) and one global score. Content analysis was reasonably supported by the theoretical foundation and the Virtual Patient expert team. The think-aloud studies and analysis of free text comments supported the validity of the instrument. In the exploratory factor analysis, using 2547 student evaluations from three different countries of a total of 78 Virtual Patients, a three-factor model showed a reasonable fit with the data. At least 200 student responses are needed to obtain a reliable evaluation of a Virtual Patient on all three factors. We conclude that the instrument has the potential to provide valid information about Virtual Patient design, provided that many responses per Virtual Patient are available.

In **Chapter 5** we search for evidence to guide the implementation of Virtual Patients. We explore students' perceptions of different scenarios aimed at developing clinical reasoning skills in which Virtual Patients were integrated with other educational activities. The analysis of eight focus group interviews suggested six themes which students considered important for the optimal implementation of VPs: (i) continuous and stable online access, (ii) increasing complexity, adapted to students' knowledge, (iii) VP-related workload offset by elimination of other activities, (iv) optimal sequencing (e.g.: lecture – 1 to 2 VP(s) – tutor-led small group discussion – real patient), (v) optimal alignment of Virtual Patients and educational activities, and (vi) inclusion of VP topics in assessment. We conclude that the themes appear to offer starting points for the development of a framework to guide the curricular integration of Virtual Patients but that their impact needs to be confirmed by further studies.

In **Chapter 6** we explore the educational impact of two different types of assessment questions on the learning of clinical reasoning with Virtual Patients. After exposure to a clinical clerkship and its exam students discussed in focus groups the perception and impact of Key-Feature Problems with Long Menu and context-rich Single Best Answer questions. Further, the psychometric characteristics of both formats were compared. The analysis of the focus groups resulted in four themes: Key-Feature Problems with Long Menu were perceived as (i) more realistic, (ii) more difficult and (iii) more motivating for the intense study of clinical reasoning with Virtual Patients than the context-rich Single Best Answer format and (iv) showed an overall good acceptance when some preconditions were taken into account. The statistical analysis revealed that there was no difference in difficulty; however, Key-Feature Problems with Long Menu showed a higher reliability (G coefficient) even when corrected for testing time. The correlations with the OSCE results were higher in the Key-Feature Problems. We conclude that students perceived the Key-Feature Problems as more motivating for the study of clinical reasoning with Virtual Patients. Including Key-Feature Problems with Long Menu into summative clerkship exams seems to offer positive educational effects without psychometric drawbacks.

Chapter 7 discusses how the previous chapters have answered the two main research questions and which conclusions and implications this yields for educational practice and further research.

Based on Chapters 2-4 the following three main aspects seem to be especially relevant for optimising VP design to foster learning of clinical reasoning: (i) Using instructional design criteria such as ensuring an appropriate level of difficulty, authenticity, interactivity, feedback and focusing on relevant learning points, (ii) implementing virtual coaching on clinical reasoning into the VP such as asking for discriminating and confirming features and (iii) providing validated instruments for systematic further improvements such as the developed VP typology and VP design questionnaire. Based on the Chapters 5-6 the following three main aspects are especially relevant for VP implementation: (iv) Sequencing VPs and other educational activities including bedside teaching with real patients according to complexity, and (v) aligning instruction and (vi) assessment with the use of VPs.

Our results are in line with current theories and insights outside of VP research e.g. on instructional design, on how to foster learning of clinical reasoning and on instructional design theories regarding curriculum development.

The main implications of our research for educational practice are that when using Virtual Patients they should be well designed, systematically improved by using validated instruments and well implemented as evidenced in this dissertation to optimally foster clinical reasoning.

Research is needed to develop a greater understanding of Virtual Patient design and implementation multicentre and in other contexts, including their longitudinal implementation and measuring their impact on patient outcomes.

Samenvatting (summary in Dutch)

Hoofdstuk 1 vormt het inleidend hoofdstuk van dit proefschrift, waarin we de wetenschappelijke achtergrond, uitgangspunten, doel en overkoepelende onderzoeksvragen beschrijven.

Door klinisch te redeneren zijn artsen in staat de juiste diagnose te stellen en over te gaan tot het juiste behandelplan voor een patiënt. Klinisch redeneren wordt doorgaans beschouwd als een bepalend onderdeel van bijna al het handelen van artsen in de praktijk. De vaardigheid van een arts om klinisch te kunnen redeneren is als zodanig van groot belang en betekenis voor het herstel en de veiligheid van patiënten. Er zijn echter tekenen die erop wijzen dat het klinisch redeneren vandaag de dag niet goed wordt onderwezen. De literatuur laat zien dat met name zogenaamde Virtuele Patiënten (VP's) geschikt zijn om het leren van klinisch redeneren te bevorderen. Virtuele patiënten zijn online gepresenteerde casussen waarbij de student zich plaatst in de schoenen van de arts en zelf de beslissingen neemt, waarop hij/zij feedback ontvangt. Zij worden gedefinieerd als "interactieve computersimulatie[s] van echte klinische scenario's ten behoeve van de Geneeskundeopleiding, het medische onderwijs of toetsing". Twee aspecten bleken van groot belang bij het leren met behulp van elektronische leermiddelen zoals VP's, namelijk: hun ontwerp en uitvoering.

Toen we aan dit onderzoek begonnen, was er maar weinig literatuur beschikbaar over het ontwerpen en uitvoeren van Virtuele Patiënten ter bevordering van het klinisch redeneren. Ons onderzoeksproject, waarmee we bewijs wilden vinden voor de verbetering van het leren met behulp van Virtuele Patiënten ter bevordering van het klinisch redeneren door Geneeskundestudenten, werd daarom ingeluid met de volgende twee onderzoeksvragen:

I: Hoe kunnen we het ontwerp van Virtuele Patiënten ter bevordering van het leren van met name klinisch redeneren verbeteren?

II: Hoe kunnen we de uitvoering van Virtuele Patiënten ter bevordering van het leren van met name klinisch redeneren verbeteren?

In **Hoofdstuk 2** gaan we na met welke aspecten er volgens studenten bij voorkeur rekening gehouden moet worden bij het ontwerpen van Virtuele Patiënten ter bevordering van hun leren, met bijzondere aandacht voor het klinisch redeneren. Nadat studenten ten minste acht in opzet variërende Virtuele Patiënten te zien hadden gekregen, bespraken zij in focusgroepen het ontwerp van Virtuele Patiënten. De studenten waren van mening dat het leren wordt bevorderd wanneer Virtuele Patiënten relevant, voldoende complex en interactief zijn, specifieke feedback geven, op de juiste manier van verschillende media gebruik maken, hen helpen focussen op relevante leerpunten, op samenvatting van de voornaamste leerpunten aansturen, echt zijn wat betreft de online interface en studenttaken en ten slotte vragen stellen en uitleg geven die het klinisch redeneren bevorderen. Nadat studenten met goed ontworpen Virtuele Patiënten hadden gewerkt, voelden zij zich terdege klaargestoomd om het klinisch redeneren op

echte patiënten toe te passen, beter dan zij met enig ander bestaande methode eerder hadden ervaren. Wij concluderen dat studenten vonden dat de beschreven basisregels voor het ontwerp hun leren bevorderden en dat vele van deze basisregels ook naar voren komen in resultaten van niet op VP's gerichte gepubliceerde studies en dat verder onderzoek nodig is.

In **Hoofdstuk 3** beschrijven we hoe wij op basis van de literatuur en een analyse van bestaande VP-systemen, en met behulp van een aangepaste Delphi-methode waarbij deskundigen op het gebied van Virtuele Patiënten tot een internationale consensus kwamen, een typologie van een Virtuele Patiënt ontwikkelden. Bij deze op empirische wijze verkregen typologie van het ontwerp van Virtuele Patiënten hingen we 19 factoren aan vier categorieën die relevant geacht werden voor het ontwerp van Virtuele Patiënten: algemeen (titel, beschrijving, taal, onderscheidende kenmerken, oorsprong, gemiddelde studietijd); onderwijskundig (onderwijsniveau, onderwijsmethodes, toepassingsgebied, doeleinden); onderwijsontwerp (type traject, gebruikersmodaliteit, gebruik van media, tekst en feedback, en mate van interactiviteit); technisch (moedersysteem, vormgeving, inbedding en afhankelijkheid). We concluderen dat deze op empirische wijze verkregen typologie van het ontwerp van Virtuele Patiënten als gemeenschappelijk referentiepunt kan dienen voor eenieder die over Virtuele Patiënten wenst te documenteren of deze wenst te onderzoeken.

In **Hoofdstuk 4** beschrijven we hoe we, met inachtneming van gepubliceerd bewijsmateriaal waaronder de lessen uit hoofdstuk 2 en 3 m.b.t. het ontwerp van Virtuele Patiënten ter bevordering van het leren van klinisch redeneren, een instrument waarmee het ontwerp van Virtuele Patiënten beoordeeld kan worden ontwikkelden en valideerden. Het korte instrument dat we verkregen bestaat uit zes vragen verdeeld over drie subschalen (echtheid van het patiëntcontact en consult, cognitieve strategieën tijdens het consult en begeleiding tijdens het consult) en een totaalscore. De inhoudsanalyse werd redelijk ondersteund door de theoretische onderbouwing en het team van deskundigen op het gebied van Virtuele Patiënten. De validiteit van het instrument werd onderzocht door de hardop-denkstudies en analyse van het in het vrije tekstveld ingevulde commentaar. Uit de exploratieve factoranalyse waarbij 2547 studenten uit drie verschillende landen in totaal 78 Virtuele Patiënten beoordeelden kwam naar voren dat een driefactorenmodel redelijk paste bij de data.

Er zijn ten minste 200 studentbeoordelingen nodig om tot een betrouwbare evaluatie van een Virtuele Patiënt te komen; dit geldt voor alle drie de factoren. We kunnen stellen dat met het instrument valide informatie verkregen kan worden over het ontwerp van Virtuele Patiënten, mits er per Virtuele Patiënt veel beoordelingen beschikbaar zijn.

In **Hoofdstuk 5** gaan we op zoek naar bewijsmateriaal dat als leidraad kan dienen voor de uitvoering van Virtuele Patiënten. Hiertoe peilen we studentwaarderingen van verschillende op de ontwikkeling van klinisch-redeneervaardigheden gerichte scenario's waarbij Virtuele Patiënten werden geïntegreerd met andere onderwijsactiviteiten. Bij de analyse van acht focusgroepinterviews kwamen zes thema's bovendrijven die studenten van belang achtten voor een optimale uitvoering van VP's: (i) een ononderbroken en stabiele online toegang, (ii) een oplopende moeilijkheidsgraad, toegesneden op de kennis van de studenten, (iii) schrappen van andere activiteiten ter compensatie van VP-gerelateerde werkbelasting, (iv) een optimale volgorde (bijv. lezing – 1 tot 2 VP[s] – discussie in kleine tutorgroepen – echte patiënt), (v) een zo goed mogelijke afstemming tussen Virtuele Patiënten en onderwijsactiviteiten, en (vi) het opnemen van VP-onderwerpen bij toetsing. We concluderen dat de thema's uitgangspunten lijken te bieden voor de ontwikkeling van een kader dat de inbedding van Virtuele Patiënten in het curriculum kan sturen, maar dat de gevolgen ervan nog door nadere studies bevestigd dienen te worden.

In **Hoofdstuk 6** onderzoeken we het leereffect van twee verschillende soorten toetsvragen op het leren klinisch redeneren met Virtuele Patiënten. Na het doen van een co-schap en bijbehorende toets bespraken studenten in focusgroepen hun ervaringen met en effecten van casusscenario's met de nadruk op essentiële beslissingen (KFP's) en uitgebreid keuzemenu (LM) enerzijds (Key-Feature Problems (KFP's) with Long Menu (LM) en contextrijke meerkeuzevragen waarbij slechts één antwoord de beste was (Single Best Answer questions) anderzijds. Verder werden de psychometrische kenmerken van beide toetsvormen met elkaar vergeleken. De analyse van de focusgroepen leidde tot vier thema's: casusscenario's met een *key-feature approach* werden beschouwd als (i) realistischer, (ii) moeilijker en als (iii) een grotere stimulans voor het intensief leren van klinisch redeneren met behulp van Virtuele Patiënten dan de contextrijke *Single Best Answer* meerkeuzevragen, en zij (iv) werden onder bepaalde voorwaarden over het algemeen goed geaccepteerd. De statistische analyse liet geen verschillen zien in moeilijkheidsgraad. Daar stond echter tegenover dat KFP's met LM een betere betrouwbaarheid (betrouwbaarheidscoëfficiënt) vertoonden, zelfs nadat gecorrigeerd was voor toetstijd. De resultaten van het stationsexamen (ook wel bekend als *Objective Structured Clinical Examinations* (OSCE's)) correleerden sterker met de KFP's. Onze conclusie is dat studenten KFP's als een grotere stimulans voor het leren van klinisch redeneren met Virtuele Patiënten beschouwden. Het opnemen van casusscenario's met de nadruk op essentiële beslissingen en uitgebreid keuzemenu in summatieve co-schapexamens lijkt te leiden tot positieve leereffecten zonder psychometrische beperkingen.

In **Hoofdstuk 7** wordt besproken hoe de voorgaande hoofdstukken een antwoord gaven op de twee onderzoeksvragen die in dit proefschrift centraal stonden, welke conclusies

we hieruit kunnen trekken en welke gevolgen dit heeft voor de onderwijspraktijk en verder onderzoek.

Op basis van hoofdstuk 2 t/m 4 lijken met name de volgende drie aspecten van belang te zijn bij het optimaliseren van het VP-ontwerp ter bevordering van het leren van klinisch redeneren: (i) toepassing van onderwijsontwerpcriteria, zoals zorgen voor de juiste moeilijkheidsgraad, mate van echtheid, interactiviteit, feedback en een focus op relevante leerpunten, (ii) zodanige programmering van de VP dat deze virtuele begeleiding biedt bij het klinisch redeneren, zoals vragen naar onderscheidende en bevestigende kenmerken, en (iii) ontwikkeling van gevalideerde instrumenten om het ontwerp stelselmatig te kunnen blijven verbeteren, zoals de ontwikkelde VP-typologie en de VP-ontwerpvragenlijst. Op basis van hoofdstuk 5 en 6 kunnen we stellen dat met name de volgende drie aspecten van belang zijn bij de uitvoering van VP's: (iv) geleidelijke opvoering van de moeilijkheidsgraad door studenten in de juiste volgorde bloot te stellen aan VP's en andere onderwijsactiviteiten waaronder praktijkonderwijs met echte patiënten (het zgn. *bedside teaching*), en (v) afstemming van het onderwijs en (vi) toetsing op het gebruik van VP's.

Onze resultaten sluiten aan bij actuele theorieën en inzichten van buiten het VP-onderzoeksgebied, bijv. vanuit de onderwijsontwerphoek, van onderzoek naar hoe het leren klinisch redeneren bevordert kan worden en naar onderwijsontwerptheorieën ten behoeve van curriculumontwikkeling. De belangrijkste gevolgen van ons onderzoek voor de onderwijspraktijk zijn dat wanneer men Virtuele Patiënten wil inzetten, men deze goed moet ontwerpen, stelselmatig moet verbeteren met behulp van gevalideerde instrumenten en moet zorgdragen voor een goede invoering zoals beschreven in dit proefschrift om het klinisch redeneren zo goed mogelijk te bevorderen. Er is onderzoek nodig om ons inzicht in het ontwerp van Virtuele Patiënten te verruimen, alsook invoering in verschillende instanties en in andere contexten, onder andere over een langere tijd genomen, zodat de uitwerking ervan op de patiëntenzorg gemeten kan worden.

REFERENCES

- Aagaard EM, Teherani A, Irby DM. Effectiveness of the one-minute preceptor model for diagnosing the patient and the learner: Proof of concept. *Academic Medicine*, 2004; 79: 42-8.
- Anderson T & Shattuck J (2012): Design-Based Research: A Decade of Progress in Education Research? *Educational Researcher* January/February 2012 vol. 41 no. 1 16-25
- Bateman J, Allen M, Samani D, Kidd J, Davies D: Virtual patient design: exploring what works and why. A grounded theory study. *Med Educ*. 2013 Jun;47(6):595-606.
- Berman N, Fall LH, Smith S, Levine DA, Maloney CG, Potts M, Siegel B, Foster-Johnson L (2009): Integration strategies for using virtual patients in clinical clerkships. *Acad Med* 84:942-949.
- Berman NB, Fall LH, Chessman AW, Dell MR, Lang VJ, Leong SL, Nixon LJ, Smith S: A collaborative model for developing and maintaining virtual patients for medical education. *Med Teach*. 2011;33:319-24.
- Cilliers FJ, Schuwirth LW, van der Vleuten CP (2012): A model of the preassessment learning effects of assessment is operational in an undergraduate clinical context. *BMC Med Educ* 12:9.
- Cook DA, Levinson AJ, Garside S (2011): Method and reporting quality in health professions education research: a systematic review. *Med Educ*. 2011 45:227-38. Review.
- Durning SJ, Dong T, Artino AR Jr, LaRochelle J, Pangaro LN, van der Vleuten C, Schuwirth L (2012): Instructional authenticity and clinical reasoning in undergraduate medical education: a 2-year, prospective, randomized trial. *Mil Med*. 2012 Sep;177(9 Suppl):38-43.
- Desjardins I, Touchie C, Pugh D, Wood TJ, Humphrey-Murto S (2014): The impact of cueing on written examinations of clinical decision making: a case study. *Med Educ*. 2014;48:255-61.
- Eva KW (2005): What every teacher needs to know about clinical reasoning. *Med Educ* 2005;39:98-106.
- Frank JR, Snell LS, Ten Cate O, Holmboe ES, Carraccio C, Swing, Harris P, Glasgow NJ, Campbell C, Dath D, Harden RM, Iobst W, Long DM, Mungroo R, Richardson DL, Sherbino J, Silver J, Taber S, Talbot M & Harris KA (2010): Competency-based medical education: theory to practice, *Medical Teacher*, 32:8, 638-645
- Frederiksen N (1984): The real test bias: Influences of testing on teaching and learning. *Am Psychol* 39:193-202.
- Higgs, J, Jones M, Loftus S, Christensen N: Clinical reasoning in the health professions. 2008: Elsevier Health Sciences.
- Huwendiek S, Hanebeck B, Bosse HM, Haag M, Hoffmann GF, Tönshoff B (2009b): Lernen und Prüfen mit virtuellen Patienten am Zentrum für Kinder- und Jugendmedizin des Universitätsklinikums Heidelberg: Ergebnisse der Evaluation im Rahmen des E-Learning-Preises Baden-Württemberg 2007. *GMS Med Inform Biom Epidemiol*. 2009;5(1):Doc10.
- Huwendiek S, Haider HR, Fischer MR, Hoffmann GF, Tönshoff B (2011): Kernausbildungsinhalte Pädiatrie für Medizinstudierende: Ergebnisse einer bundesweiten Umfrage unter Lehrbeauftragten der Universitätskinderkliniken und Obmännern des Berufsverbands der Kinder- und Jugendärzte e.V. *Monatsschr Kinderheilkd* 2011; 159:655-661
- Huwendiek S, Klinke M, Lehmann R, Simon A, Tönshoff B, Meyburg J (2012): Ideal design of a postgraduate interprofessional blended learning concept to improve paediatric emergency care: results of a focus group study among participants and tutors. *Kongress der Association for Medical Education in Europe*, 2012, Lyon, France, abstract 3Y/14, p. 106
- Kern DE, Thomas PA, Hughes MT (Eds) 2010: Curriculum Development for Medical Education: A Six-Step Approach. Johns Hopkins University Press
- Kogan, J. R., Holmboe, E. S., & Hauer, K. E. (2009). Tools for direct observation and assessment of clinical skills of medical trainees: a systematic review. *Jama*, 302(12), 1316-1326.
- Lehmann R, Thiessen C, Frick B, Bosse HM, Nikendei C, Hoffmann GF, Tönshoff B, Huwendiek S: Improving Pediatric Basic Life Support Performance Through Blended Learning With Web-Based Virtual Patients: Randomized Controlled Trial. *J Med Internet Res*. 2015 Jul 2;17(7):e162.
- Mamede S, van Gog T, Sampaio AM, de Faria RMD, Maria JP, Schmidt HG (2014): How can students' diagnostic competence benefit most from practice with clinical cases? The effects of structured reflection on future diagnosis of the same and novel diseases. *Acad Med* 2014;89 (1):121-7.

- McGaghie WC, Issenberg SB, Cohen ER, Barsuk JH, Wayne DB (2011): Does Simulation-based Medical Education with Deliberate Practice Yield Better Results than Traditional Clinical Education? A Meta-Analytic Comparative Review of the Evidence. *Acad Med.* 2011 Jun; 86(6): 706–711.
- Molenda M (2003): In Search of the Elusive ADDIE Model. *Performance Improvement*, 42(5), 34-36.
- Norcini JJ, Boulet JR, Opalek A, Dauphinee WD (2014): The relationship between licensing examination performance and the outcomes of care by international medical school graduates. *Acad Med.* 2014 Aug;89(8):1157-62. Do
- Norman G: Teaching basic science to optimize transfer. *Medical teacher* 2009, 31(9):807-811.
- Norman G. Research in clinical reasoning: past history and current trends. *Med Educ* 2005;39:418–27.
- O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA (2014): Standards for reporting qualitative research: a synthesis of recommendations. *Acad Med.* 2014 Sep;89(9):1245-51.
- Posel N, McGee JB, Fleiszer DM (2015): Twelve tips to support the development of clinical reasoning skills using virtual patient cases. *Med Teach.* 2015 Sep;37(9):813-8.
- Prober CG & Heath C (2012): Lecture halls without lectures—a proposal for medical education. *N Engl J Med.* 2012 May 3;366(18):1657-9
- Prober CG & Khan S (2013): Medical education reimaged: a call to action. *Acad Med.* 2013 Oct;88(10):1407-10.
- Rattner SL, Louis DZ, Rabinowitz C, Gottlieb JE, Nasca TJ, Markham FW, Gottlieb RP, Caruso JW, Lane JL, Veloski J, Hojat M, Gonnella JS. Documenting and medical students' comparing clinical experiences. *JAMA* 2001;286 (9):1035–40
- Schmidt HG & Mamede S (2015): How to improve the teaching of clinical reasoning: a narrative review and a proposal. *Med Educ.* 2015 Oct;49(10):961-73. van de Ridder JM, McGaghie WC, Stokking KM, ten Cate OT (2015): Variables that affect the process and outcome of feedback, relevant for medical training: a meta-review. *Med Educ.* 2015 Jul;49(7):658-73. doi: 10.1111/medu.12744.
- Schifferdecker KE, Berman NB, Fall LH, Fischer MR. Adoption of computer-assisted learning in medical education: the educators' perspective. *Med Educ.* 2012 Nov;46:1063-73.
- SNAPPS (Wolpaw T, Papp KK and Bordage G. Using SNAPPS to facilitate the expression of clinical reasoning and uncertainties: a randomized comparison group trial. *Academic Medicine* 2009;84(4):517-24.
- Swanson DB, Holtzman KZ, Clauser BE, Sawhill AJ: Psychometric characteristics and response times for one-best-answer questions in relation to number and source of options. *Acad Med.* 2005 Oct;80(10 Suppl):S93-6.
- Thompson L, Gentner D, Loewenstein J (2000): Avoiding missed opportunities in managerial life: analogical learning improves case-based transfer. *Organ Behav Human Decis Process* 2000;47:98–123.
- Wayne DB, Butter J, Siddall VJ, Fudala MJ, Wade LD, Feinglass J, McGaghie WC. Mastery learning of advanced cardiac life support skills by internal medicine residents using simulation technology and deliberate practice. *J Gen Intern Med.* 2006;21:251-6.
- Williams RG, Klamen DL, White CB, Petrusa E, Rincher RME, Whitfield CF, Shatzer JH, McCarty T, Miller BM (2011): Tracking development of clinical reasoning ability across five medical schools using a progress test. *Academic Medicine.* 2011; 86(9): 1148-1154.
- Wimmers PF, Schmidt HG, Splinter TAW. Influence of clerkship experiences on clinical competence. *Med Educ* 2006;40:450–8.
- Yudkowsky R, Otaki J, Lowenstein T, Riddle J, Nishigori H, Bordage G: A hypothesis-driven physical examination learning and assessment procedure for medical student: initial validity evidence. *Med Educ* 2009;43:729-40.

Valorization

SOCIAL RELEVANCE

The research reported in this dissertation is not only about medical education but also about patient care and patient safety issues. Medical education and health care delivery are directly interconnected in the projects described. As outlined in the introduction, clinical reasoning is a central component of nearly everything doctors do in practice. Clinical reasoning enables doctors to establish the right diagnosis and make the right treatment decisions for a patient. Therefore, the research in this dissertation, which deals with facilitating learning of clinical reasoning, is of societal relevance.

As three of the five studies (Chapters 2, 5, 6) were conducted within a real setting with medical students seeing real patients, there is already a direct impact of the projects and research on society.

ECONOMIC RELEVANCE

Diagnostic failures do not only have a big societal impact on patients, they also have economic implications. If medical students and doctors have better clinical reasoning skills, e.g., when learning is additionally facilitated by VPs, this can have a bearing on economic measures; however, we did not measure this in the present dissertation.

Clinical supervisors' time can be used more efficiently if students have practised with VPs before seeing real patients. It is very much worth the effort developing VPs if many students will use them, as they allow supervisors to focus on the provision of feedback on more difficult aspects and help students apply their skills to the individual real patient, rather than introducing the basics.

Developing well-designed VPs is a laborious task, and it should therefore be considered to develop them in collaboration with multiple centres as outlined in the 'implications for future research' section of this chapter. In such cases, less money and effort may be needed to provide well-designed VPs. This issue has been addressed in successful projects where also the main author of this dissertation was involved (eVIP: <http://virtualpatients.eu/>).

TARGET GROUPS

The main rationale for this research was the need to improve clinical reasoning education of medical students by means of Virtual Patients. The target groups of this intervention included medical students (as the end users) and other stakeholders such as medical educators, medical teachers, supervisors and curriculum planners.

Supervision and feedback are central aspects in almost every educational programme. Therefore, the lessons learned from this dissertation could also be helpful in other settings.

INNOVATION

This dissertation is innovative in that it presents new ways to design, implement, report on and evaluate Virtual Patients. While there are others working on the design and implementation of VPs, the publication of the described results in respected journals in the field also shows that these results are seen as innovative.

SCHEDULE AND IMPLEMENTATION

As mentioned in Chapters two, five and six, VPs are already being used extensively in Heidelberg Medical School. In collaboration projects they have also been used in other places and there are plans to extend their use (after repurposing) to other German, Swiss and international faculties, as well as to other levels such as postgraduate education. Usually, external users must pay a support fee to use the software. The cases themselves are usually not sold per se but rather exchanged in collaboration projects. For the last several years, I have been discussing the potential of VPs at academic conferences and asking other researchers for their reactions to them. This has been of immense help and led to invitations to join other collaborations on VPs.

ACTIVITIES/PRODUCTS

The primary products of this dissertation are the published papers it comprises. The described typology (Chapter 3) and questionnaire (Chapter 4) can be and have already been used by other researchers and VP developers. The VPs are presently used in Heidelberg and other universities. The following pages will give an overview of activities concerning VPs, such as publications, invited presentations and workshops, which are not part of this PhD dissertation itself.

Published articles concerning VPs besides the ones for this PhD

Berman NB, Durning SJ, Fischer MR, Huwendiek S, Triola MM (2016): The Role for Virtual Patients in the Future of Medical Education. Acad Med. 2016 Mar 8. [Epub ahead of print]

- Barteit S, Hoepffner P, Huwendiek S, Karamagi A, Munthali C, Theurer A, Neuhaus F. Self-directed e-learning at a tertiary hospital in Malawi – A qualitative Evaluation and Lessons learnt. *GMS Z Med Ausbild.* 2015;32(1):Doc7.
- Lehmann R, Hanebeck B, Oberle S, Simon A, Choukair D, Tönshoff B, Huwendiek S. Virtual patients in continuing medical education and residency training: a pilot project for acceptance analysis in the framework of a residency revision course in pediatrics. *GMS Z Med Ausbild.* 2015;32(5):Doc51.
- Lehmann R, Thiessen C, Frick B, Bosse HM, Nikendei C, Hoffmann GF, Tönshoff B, Huwendiek S: Improving Pediatric Basic Life Support Performance Through Blended Learning With Web-Based Virtual Patients: Randomized Controlled Trial. *J Med Internet Res* 2015;17(7):e162
- Lehmann R, Bosse HM, Simon A, Nikendei C, Huwendiek S. An innovative blended learning approach using virtual patients as preparation for skills laboratory training: perceptions of students and tutors. *BMC Med Educ.* 2013 Feb 12;13(1):23.
- Huwendiek S, Lehmann R, Simon A, Hoffmann GF, Tönshoff B: Einsatz innovativer Lehrmethoden und Herausforderungen in der pädiatrischen universitären Lehre drei Jahre nach Änderung der Approbationsordnung: Ergebnisse einer deutschland-weiten Umfrage. *Monatsschr Kinderheilkd* 2012; 160: 260–266
- Triola MM, Huwendiek S, Levinson AJ, Cook DA: New Directions in e-Learning Research in Health Professions Education: Report of Two Symposia, *Med Teach.* 2012;34(1): e15-20.
- Huwendiek S, Haider HR, Fischer MR, Hoffmann GF, Tönshoff B: Kernausbildungsinhalte Pädiatrie für Medizinstudierende: Ergebnisse einer bundesweiten Umfrage unter Lehrbeauftragten der Universitätskinderkliniken und Obmännern des Berufsverbands der Kinder- und Jugendärzte e.V. *Monatsschr Kinderheilkd* 2011; 159:655–661
- Oberle S, Huber S, Nawrotzki R, Harter C, Schönit E, Schäfer N, Tönshoff B, Huwendiek S: Evaluations of different designs of virtual patients: First results using the eVIP evaluation instruments. *Bio-Algorithms and Med-Systems* 2010; 6: 15-19
- Davies D, Balasubramaniam C, Huwendiek S, De Leng B: Needs analysis for Virtual Patients: a report from the eVIP project. *Bio-Algorithms and Med-Systems* 2009; 5: 29-34
- Balasubramaniam C, Poulton T, Huwendiek S: Repurposing existing virtual patients; an anglo-german case study. *Bio-Algorithms and Med-Systems* 2009; 5: 91-98
- Huwendiek S, Haider HR, Tönshoff B, De Leng B: Evaluation of curricular integration of virtual patients: Development of a student questionnaire and a reviewer checklist within the electronic virtual patient (eVIP) project. *Bio-Algorithms and Med-Systems* 2009; 5: 35-44
- Hanebeck B, Tönshoff B, Huwendiek S: Repurposing virtual patients for clinical reasoning: Development of a guideline and assessment of time and effort. *Bio-Algorithms and Med-Systems* 2009; 5: 69-72
- Huber S, Hess F, Donkers J, Heid J, Huwendiek S, De Leng B, Haag M: Enriching Virtual Patients with Interactive Images. *Bio-Algorithms and Med-Systems* 2009; 5: 73-77

- Oberle S, Huber S, Tönshoff B, Nawrotzki R, Huwendiek S: Repurposing Virtual Patients for the preclinical years - A Pilot study. *Bio-Algorithms and Med-Systems* 2009; 5: 79-82
- Hess F, Heid J, Huber S, Huwendiek S, Haag M: Standardkonforme systemübergreifende Nutzung virtueller Patienten durch Umsetzung des MedBiquitous Virtual Patient Standards am Beispiel von CAMPUS. *GMS Med Inform Biom Epidemiol.* 2009;5(1): Doc08.
- Huwendiek S, Hanebeck B, Bosse HM, Haag M, Hoffmann GF, Tönshoff B: Lernen und Prüfen mit virtuellen Patienten am Zentrum für Kinder- und Jugendmedizin des Universitätsklinikums Heidelberg: Ergebnisse der Evaluation im Rahmen des E-Learning-Preises Baden-Württemberg 2007. *GMS Med Inform Biom Epidemiol.* 2009;5(1): Doc10.
- Huwendiek S, Kadmon M, Jünger J, Kirschfink M, Bosse HM, Resch F, Duelli R, Bardenheuer HJ, Sonntag HG, Steiner T: Umsetzung der deutschen Approbationsordnung 2002 im modularen Reformstudiengang Heidelberger Curriculum Medicinale (Hei-CuMed). *ZFHE Zeitschrift für Hochschulentwicklung*, 3, 17-27, 2008. Verfügbar unter: <http://www.zfhe.at>
- Huwendiek S, Hanebeck B, Bosse HM, Haag M, Hoffmann GF, Tönshoff B: Lernen und Prüfen mit Virtuellen Patienten am Zentrum für Kinder- und Jugendmedizin Heidelberg. 12. Workshop der Arbeitsgruppe "Computerunterstützte Lehr- und Lernsysteme in der Medizin" der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie, Proceedings, 17-20, 2008
- Bauch M, Hess F, Heid J, Leven FJ, Huwendiek S, Haag M: Verwendung eines standardisierten Austauschformats - Implikationen für existierende Lernsysteme am Beispiel der CAMPUS-Shell. 12. Workshop der Arbeitsgruppe "Computerunterstützte Lehr- und Lernsysteme in der Medizin" der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie, Proceedings, 27-29, 2008
- Fischer MR, Hege I, Puppe F, Tönshoff B & Huwendiek S: Virtuelle Patienten in der medizinischen Ausbildung: Vergleich verschiedener Strategien zur curricularen Integration. *Z. Evid. Fortbild. Qual. Gesundh.wesen (ZEFAQ)* 102, 648-653, 2008
- Reichert F, Brasch C, Heid J, Bosse HM, Haag M, Leven FJ, Hoffmann GF, Tönshoff B, Huwendiek S: Design virtueller Patienten für die medizinische Ausbildung: Evaluation zweier Abspielvarianten des CAMPUS-Systems durch Medizinstudierende. *eLearning in der Medizin und Zahnmedizin*, Hrsg.: Kundt, Bernauer, Fischer, Haag, Klar, Leven, Matthies, Puppe, Shaker Verlag: 222 - 234, 2007
- Garde S, Bauch M, Haag M, Heid J, Huwendiek S, Ruderich F, Singer R, Leven FJ: CAMPUS - Computer-Based Training in Medicine as part of a Problem-oriented Educational Strategy. *Studies in Learning, Evaluation, Innovation and Development* 2(1), 10-19, 2005. Abrufbar unter: <http://sleid.cqu.edu.au/viewissue.php?id=6>

Short communications

- Huwendiek S & de Leng BA: Virtual patient design and curricular integration evaluation toolkit. Med Educ. 2010;44:519.
- Lehmann R, Bosse HM, Huwendiek S: Blended learning using virtual patients and skills laboratory training. Med Educ. 2010 44:521-2
- Huwendiek S, Köpf S, Höcker B, Heid J, Bauch M, Bosse HM, Haag M, Leven FJ, Hoffmann GF, Tönshoff B: Fünf Jahre Erfahrung mit dem curricularen Einsatz des fall- und webbasierten Lernsystems "CAMPUS-Pädiatrie" an der Medizinischen Fakultät Heidelberg. GMS Z Med Ausbild. 23: Doc10, 2006

Review articles

- Haag M & Huwendiek S: The virtual patient for education and training: a critical review of the literature. It- Information Technology 2010, 52: 281-287
- Huwendiek S, Muntau AC, Maier EM, Tönshoff B, Sostmann K: E-Learning in der medizinischen Ausbildung: Ein Leitfaden zum sinnvollen Einsatz. Monatsschr. Kinderheilkunde 156: 458-463, 2008

Book articles

- Huwendiek s, Haag M: Der virtuelle Patient im Rahmen der medizinischen Ausbildung. In: Niederlag W, Lemke HU, Lehrach H, Peitgen H-O (Hrsg.): Der Virtuelle Patient, Health Academy, De Gruyter, 2. erweiterte Auflage 2014, S.207-216
- Huwendiek S & Haag M: Der virtuelle Patient im Rahmen der medizinischen Ausbildung. In: Niederlag W, Lemke HU, Lehrach H, Peitgen H-O (Hrsg.): Der Virtuelle Patient, Health Academy, Dresden 2012, S. 229-239
- Huwendiek S, Köpf S, Heid J, Höcker B, Bosse HM, Haag M, Leven FJ, Hoffmann GF, Tönshoff B: CAMPUS-Pädiatrie: Fallbasierte Aus- und Weiterbildung in der Kinderheilkunde und Jugendmedizin. In: Lehmann/Bloh (Hrsg.). Online-Pädagogik Band 3: Referenzmodelle und Praxisbeispiele, Schneider Verlag Hohengehren, Baltmannsweiler: 133 - 143, 2005
- Huwendiek S, Köpf S, Singer R, Höcker B, Hoffmann GF, Leven FJ, Tönshoff B: CAMPUS-Pädiatrie: Etablierung von fall- und webbasiertem Training in der Kinderheilkunde und Jugendmedizin. In: Christoph Brake, Monika Topper, Joachim Wedekind (Hrsg.), "Der MEDIDA-PRIX - Nachhaltigkeit durch Wettbewerb", Reihe Medien in der Wissenschaft, Band 31, Waxmann Verlag, Münster: 91 - 102, 2004
- Köpf S, Huwendiek S, Singer R, Höcker B, Hoffmann GF, Leven FJ, Tönshoff B: CAMPUS-Pädiatrie - Ein web-basiertes, interaktives, fallbasiertes Computer-Lehr- und Lernprogramm für die Ausbildung in der Kinderheilkunde. In: Hohenstein A; Wilbers K (Hrsg.):

- Handbuch E-Learning- Expertenwissen aus Wissenschaft und Praxis, 7. Erg.-Lfg. Januar 2004, Wolters Kluwer Deutschland GmbH, Neuwied, Köln, München, 2003
- Singer R, Köpf S, Martsfeld I, Heid J, Bauch M, Ruderich F, Huwendiek S, Tönshoff B, Leven FJ: CAMPUS-Pädiatrie interaktiv: Entwicklung einer CD-ROM mit interaktiven multimedialen Kasuistiken aus der Pädiatrie und integriertem LehrbuchIn: Puppe, F; Albert, J; Fischer, M; Klar, R; Leven, J (Hrsg.): Rechnergestützte Lehr- und Lernsysteme in der Medizin: Proceedings zum 7. Workshop der GMDS AG Computergestützte Lehr/- und Lernsysteme in der Medizin, Universität Würzburg, Shaker, 2003

Presentations and abstracts concerning VPs

More than 100

Invited talks concerning VPs

- Huwendiek S: Virtual Patients in Medical Education: Delegation from Mongolei. Institute of Medical Education, Bern, intern, 2015
- Huwendiek S: Blended Learning, Tag der Lehre der Uni Bern, 2015
- Huwendiek S, Vortrag: E-Learning. Seminar at the Medical Education Symposium of the Schweizerischen Instituts für Weiter- und Fortbildung, Bern, 2015
- Huwendiek S: Virtuelle Patienten. E-Learning Workshop GMA-Tagung, Leipzig, 2015.
- Huwendiek S et al. Interprofessionelles Blended Learning Weiterbildung, Kongress der Deutschen Gesellschaft für Kinder- und Jugendmedizin, 2013
- Huwendiek S: Virtuelle Patienten. Kinderklinik Bern Ausschuss Lehre, Februar 2013
- Huwendiek S: E-Learning an der medizinischen Fakultät Heidelberg. Ärztekammer Baden-Württemberg, Stuttgart, 14.3.2012
- Huwendiek S, Lehmann R, Brass C, Reichert F, de Leng BA, Dolmans D, van der Vleuten CPM, Haag M, Hoffmann GF, Tönshoff B: Comparison of summative assessment with electronic key feature cases using the long menu approach and case based multiple choice questions: a focus group study. Medical Assessment Alliance: International Conference in competency-based Assessment, Heidelberg 2012, conference abstract book p.15
- Huwendiek S: Der Virtuelle Patient in der Medizinischen Ausbildung. 18. Dresdner Palais-Gespräch: Der virtuelle Mensch – Zukünftige Basis für Prävention, Diagnose und Therapie? Dresden, 2. September 2011
- Huwendiek S: Neue Lehrformen in der Mediziner Ausbildung an der Universitätsklinik Heidelberg. Fachtagung Professional E-Learning 2011 im Rahmen der Didacta, Stuttgart, 24.2.2011
- Huwendiek S: Spotlight „E-Learning“. Plenary session on the last day of the congress of the Association for Medical Education in Europe (AMEE), Glasgow, UK, 2010

- Huwendiek S: E-Learning und E-Prüfung mit Virtuellen Patienten in der Medizin. Elektronische Prüfungen in Hochschulen – Rahmenbedingungen und Praxiserfahrungen. HIS Workshop am 05. und 06. November 2008 in Hannover
- Huwendiek S: Lernen und Prüfen mit Virtuellen Patienten. IMC Learning World: "Shaping the Future of Learning: Was bringt die Web 2.0 Innovationswelle?", Berlin, 2007
- Huwendiek S: CAMPUS and E-Learning. Delegation der Med. Faculty, Srinakharinwirot University, Bangkok, Thailand, in Heidelberg, 2007
- Huwendiek S, Bosse HM, Köpf S, Singer R, Höcker B, Hoffmann GF, Leven FJ, Tönshoff B: CAMPUS-Pediatrics and the CAMPUS assessment system. Presentation for the curriculum commission, Medical Faculty, University Ghent, Belgium, in Heidelberg, 2006
- Huwendiek S, Köpf S, Singer R, Höcker B, Hoffmann GF, Leven FJ, Tönshoff B: „CAMPUS and CAMPUS-Pediatrics“. Presentation for the curriculum commission, Medical Faculty, University of Ljubljana, Slovenia, Heidelberg, 2004
- Huwendiek S, Köpf S, Singer R, Höcker B, Hoffmann GF, Leven FJ, Tönshoff B: "CAMPUS-Pädiatrie". Learntec, Präsentation im Rahmen der Virtuellen Hochschule Baden-Württemberg, Karlsruhe, 2004
- Huwendiek S, Köpf S, Singer R, Höcker B, Hoffmann GF, Leven FJ, Tönshoff B: „CAMPUS-Pädiatrie“. Multimedia-Forum Rheinland Pfalz, Von der Illusion zu „best practice“ in E-Learning, Kaiserslautern, 10.-11.11.2003
- Huwendiek S, Köpf S, Singer R, Höcker B, Hoffmann GF, Leven FJ, Tönshoff B: "CAMPUS-Pädiatrie". Learntec, Karlsruhe, 2003
- Huwendiek S, Köpf S, Singer R, Höcker B, Hoffmann GF, Leven FJ, Tönshoff B: „CAMPUS-Pädiatrie“. 3. Workshop E-Learning der Universität Heidelberg, Heidelberg, 2003
- Huwendiek S, Köpf S, Singer R, Höcker B, Hoffmann GF, Leven FJ, Tönshoff B: „CAMPUS-Pediatrics“. Presentation for the curriculum commission of the Universidad Autonoma del Nuevo Leon, Monterrey, Mexiko, Heidelberg, 2003
- Huwendiek S, Köpf S, Singer R, Höcker B, Hoffmann GF, Leven FJ, Tönshoff B: „CAMPUS-Pediatrics“. Presentation for four postdocs from the University of Sarajevo, School of Medicine, Bosnien-Herzegowina, Heidelberg, 2003

Conducted workshops

- Haag M, Igel C, Tolks D, Baldeweg P, Graf N, Sostmann K, Huwendiek S, Dietrich M, Ulrich C, Sarikas A, Kühbeck F: eLearning in der Medizin: Best Practices und Trends. In: Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA). Hamburg, 25.-27.09.2014. Düsseldorf: German Medical Science GMS Publishing House; 2014. DocW02.
- Kirschfink M, Huwendiek S, Schulz JH, Bosse HM, Strittmatter V: Sino-German Tongji Med Workshop V, Wuhan, 15.-23.6.2012
- Huwendiek S et al.: Interprofessionelles Notfalltraining, Zentrum für Kinder- und Jugendmedizin Heidelberg, whole day (7 times in 2012)

- ...Huwendiek S...: Klausurwoche Dozentenschulung II der Medizinischen Fakultät Heidelberg, Bad Herrenalb, März 2012
- Kirschfink M, Huwendiek S, Roos M, Strittmatter V: Sino-German Tongji Medical Education Workshop. Tongji Medical School, Wuhan, China, 2011
- ...Huwendiek S...: Klausurwoche Dozentenschulung II der Medizinischen Fakultät Heidelberg Bad Herrenalb, 2011
- ...Huwendiek S...: Klausurwoche Dozentenschulung II der Medizinischen Fakultät Heidelberg, Bad Herrenalb, 2010
- ...Huwendiek S...: Klausurwoche Dozentenschulung II der Medizinischen Fakultät Heidelberg, Bad Herrenalb, 2009
- Huwendiek S & Hess F: Workshop on Virtual Patients, EU-Projekt Intel M Sarajevo, 2008
- ...Huwendiek S...: Klausurwoche Dozentenschulung II der Medizinischen Fakultät Heidelberg, Bad Herrenalb, 2008
- ...Huwendiek S...: Seminarblock Dozentenschulung I der Medizinischen Fakultät Heidelberg, Bad Herrenalb, 2008
- Huwendiek S & Bauch M: Workshop on Virtual Patients. EU-Projekt Intel M, Neum, Bosnien-Herzegowina, 2007
- ...Huwendiek S...: Workshop Train the Trainer. EU-Projekt Intel M, Neum, Bosnien-Herzegowina, 2007
- ...Huwendiek S...: Klausurwoche Dozentenschulung II der Medizinischen Fakultät Heidelberg, Bad Herrenalb, 2007
- ...Huwendiek S...: Seminarblock Dozentenschulung I der Medizinischen Fakultät Heidelberg, Bad Herrenalb, 2007
- ...Huwendiek S...: Faculty development workshop week in Bosnia Herzegowina, Jahorina (EU-Projekt Intel M), 2006
- ...Huwendiek S...: Faculty development workshop week in Bosnia Herzegowina, Neum (EU-Projekt Intel M), 2007
- Fischer M & Huwendiek S: Virtual Patients online: How to integrate them successfully in your curriculum. Workshop Kongress der Association for Medical Education in Europe (AMEE), Abstract 10M, Abstractband S. 213, 2006
- Fischer M & Huwendiek S: e-Learning. Master of Medical Education Programm Deutschland, Heidelberg, 26.7.2005
- ...Huwendiek S...: Dozententraining Fürth der medizinischen Fakultät Heidelberg Fürth, Odenwald, März 2005
- ...Huwendiek S...: Dozententraining des Instituts für Didaktik und Forschung in der medizinischen Lehre e.V.. Fürth, Odenwald, Oktober 2005

Acknowledgements

I wish to thank my three supervisors in Maastricht: Diana Dolmans, Cees van der Vleuten and Bas de Leng. You provided me with excellent guidance, very helpful and timely feedback and encouragement. It has been a great privilege working with you. I also wish to express my gratitude to Bas and Diana for their great hospitality, making the visits to Maastricht unforgettable.

Many thanks also go to my two supervisors from University Children's Hospital Heidelberg, Burkhard Tönshoff and Georg Friedrich Hoffmann, who supported me especially in my early research steps and in obtaining funding for the Master of Health Professions Education programme and this distance PhD programme, for example. I also thank Burkhard Tönshoff for being my mentor during my time at Children's Hospital Heidelberg.

I thank Martin Haag, Frank Hess, Jörn Heid and Simone Huber from Heilbronn University for the great and fruitful collaboration on Virtual Patients. You developed the nice software. Without you this dissertation would not have been possible.

I thank Sissel Guttormsen, my director at the Institute of Medical Education in Bern, for her support of my visits to Maastricht in 2015, which helped me a lot in making progress during the final writing stages of this dissertation.

I thank Friedrich Reichert, Cecilia Dunker and Rani Haider, of whom I was the MD thesis supervisor, for their great help in administering and analysing the focus groups described in this dissertation.

I also wish to thank Arno Muijtjens from Maastricht for his great statistical support regarding Chapters 4 and 6 and Daniel Stricker from the Institute of Medical Education in Bern for his statistical support concerning chapter 6.

I thank my colleagues from the European eVIP project for the inspiring collaborations across borders resulting in the research reported in Chapters 3 and 4.

I thank my colleagues Hans Martin Bosse, Ronny Lehmann, Anke Seitz, Benjamin Hanebeck, Barbara Frick and Thomas Lutz for the great collaboration on many educational innovations at the University Children's Hospital Heidelberg.

I thank the secretaries of the paediatric clerkship in Heidelberg, Sandra Telcher and Claudia Meyer, for their great support.

I thank the colleagues from the competence assessment centre Heidelberg for their support in running the electronic exams.

I thank all the students from Heidelberg Medical School and abroad for their willingness to share their experiences with Virtual Patients.

I thank Mereke Gorsira and Angelique van den Heuvel for their excellent help in the English correction of the articles and this dissertation.

I further wish to thank my brothers Carsten and Frederic and their families for the great times we have spent together and for being there when I need you.

I thank my parents Erika and Volker for their great loving support throughout my whole life.

I thank my son Elias for smiling, singing and the loads of fun we have together.

ACKNOWLEDGEMENTS

Last but not least, I thank my wife, Susanne. You have patiently tolerated much to allow me to pursue this PhD. I could never have done this without your love.

Curriculum Vitae

Sören was born on the 20th February 1974 in Mannheim, Germany.

After high school, he completed his civilian service as a paramedic and studied medicine at Heidelberg Medical School (Germany), graduating in 2001. During his medical studies, he also gained experience with studying and electives abroad, incl. St Bartholomew's and the Royal London School of Medicine (Great Britain), Karolinska Institute (Sweden), Maastricht University (the Netherlands) and Harvard Medical School (USA). His great interest in medical education already emerged during his student days, e.g. introducing Problem-Based Learning at his home faculty together with peers.

For his MD Thesis (Dr. med.) on the genomic instability of breast adenocarcinomas, he conducted biomolecular research at the German Cancer Research Center (DKFZ, Heidelberg) in cooperation with the Cancer Center Karolinska, Karolinska Institute, Sweden.

From 2002 to 2012, he completed his residency and specialization as Pediatrician and Pediatric Rheumatologist at the University Children's Hospital Heidelberg, and additionally worked as a medical educator at the Children's Hospital and Heidelberg Medical Faculty.

He finished his postgraduate Master of Medical Education in Bern in 2008, for which he received a grant from Heidelberg Medical School. Around the same time, he began his distance PhD in Medical Education at the School of Health Profession Education (Maastricht University) on Virtual Patients.

Since 2012, he has been Head of the Department of Assessment and Evaluation at the Institute of Medical Education in Bern, Switzerland. Additionally, he still sees patients in the pediatric rheumatology outpatient clinic of the University Children's Hospital Bern.

He received his senior lecture qualification (Habilitation) in Pediatrics on "Innovative Pediatric Teaching and Assessment" from Heidelberg Medical Faculty in 2014, and his senior lecture qualification in Medical Education (Umhabilitation) from Bern Medical Faculty in 2015.

He has been and is involved in several educational (inter)national projects including funding from the European commission (e.g. on Virtual Patients, eVIP) and the Swiss Federal Office of Public Health (e.g. on improving the assessment of communication skills within the Swiss Federal Licensing Exam).

He has a broad interest in teaching and research in Medical Education. Among his favorite themes are formative assessment (Virtual Patients, workplace-based assessments), innovative ways of summative assessment, communication and practical skills, Blended Learning, Problem-based learning and Medical Educators. He has conducted over 48 educational workshops locally, nationally and internationally (e.g. Association for Medical Education in Europe, China, Bosnia-Herzegovina).

He has supervised and supervises PhD, Master of Medical Education, and MD theses and is a member of several (inter)national medical education committees.

He is a reviewer of more than 12 journals and member of the editorial board of the BMC Medical Education and the GMS Journal for Medical Education. He has published more than 75 peer-reviewed manuscripts and won several medical education awards.

Sören is married to Susanne and has one son, Elias.

She dissertation series

SHE DISSERTATIONS SERIES

The SHE Dissertation Series publishes dissertations of PhD candidates from the School of Health Professions Education (SHE) who defended their PhD theses at Maastricht University. The most recent ones are listed below. For more information go to: <https://she.mumc.maastrichtuniversity.nl>

Ginsburg, S. (01-09-2016) Hidden in plain sight, the untapped potential of written assessment comments

Koops, W. (08-06-2016) Computer-supported collaborative learning in clinical clerkships

Schlegel, C. (08-06-2016) Simulated and standardized patients in health profession education: the impact of quality improvement

Sorensen, J. 01-06-2016 Obstetric simulation: designing simulation-based medical education and the role of physical fidelity

Kok, E. (01-04-2016) Developing visual expertise: from shades of grey to diagnostic reasoning in radiology

Van den Eertwegh, V. (11-11-2015) Unravelling postgraduate communication learning; from transfer to transformative learning

Gingerich, A. (03-09-2015) Questioning the rater idiosyncrasy explanation for error variance, by searching for multiple signals within the noise

Goldszmidt, M. (02-09-2015) Communication and reasoning on clinical teaching teams, the genres that shape care and education

Slootweg, I. (19-06-2015) Teamwork of Clinical Teachers in Postgraduate Medical Training

Al-Eraky, M. (21.05.15) Faculty development for medical professionalism in an Arabian context

Wearne, S. (08-04-2015) Is it remotely possible? Remote supervision of general practice registrars

Embo, M. (13-03-2015) Integrating workplace learning, assessment and supervision in health care education

Zwanikken, P. (23-01-2015) Public health and international health educational programmes for low- and middle-income countries: questioning their outcomes and impact

Hill, E. (11-12-2014) A cutting culture: gender and identification in the figured world of surgery

Diemers, A. (03-10-2014) Learning from pre-clinical patient contacts

Tijam, I. (17-09-2014) Learning in Urology. Designing simulator based skills Training & Assessment

Berkenbosch, L. (30-06-2014) Management and leadership education for medical residents

Bergman, E.M. (30-06-2014) Dissecting anatomy education in the medical curriculum

Dijkstra, J. (25-06-2014) Guidelines for designing programmes of assessment

Van Loon, M.H. (08-05-2014) Fostering monitoring and regulation of learning

Frambach, J.M. (26-03-2014) The cultural complexity of problem-based learning across the world

Hommes, J.E. (26-02-2014) How relations, time & size matter in medical education

Van der Zwet, J. (30-01-2014) Identity, Interaction and Power. Explaining the affordances of doctor-student interaction during clerkships

Watling, C.J. (22-01-2014) Cognition, Culture, and Credibility. Deconstructing Feedback in Medical Education

